University of Alberta

System Identification and Control of Multirate Systems

by

Dongguang Li

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

in

Process Control

Department of Chemical and Materials Engineering

Edmonton, Alberta
Fall 2001

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-68962-X

Canada

University of Alberta

Library Release Form

Name of Author: Dongguang Li

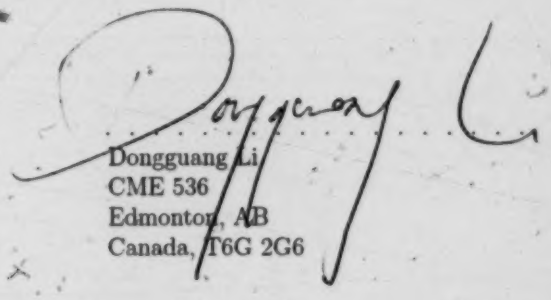
Title of Thesis: System Identification and Control of Multirate Systems

Degree: Doctor of Philosophy

Year this Degree Granted: 2001

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.



Dongguang Li
CME 536
Edmonton, AB
Canada, T6G 2G6

Date: August 31, 2001

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **System Identification and Control of Multirate Systems** submitted by Dongguang Li in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in *Process Control.*

Dr. S.L. Shah(Supervisor)

Dr. T. Chen(Co-supervisor)

Dr. B. Huang

Dr. S. Meadows

Dr. H. Marquez

for Dr. J.H. Lee (External Examiner)

Date: Sept. 12, 2001

To my wife and my parents

Abstract

Model-based predictive control (MPC) emerged in the late 1970's and has achieved tremendous success in industry. Multirate systems such as distillation columns, reactors, etc. are very common in chemical industry, but it is difficult to apply MPC to control such multirate systems since it is difficult to extract process information from multirate data. Multirate model-based inferential control schemes have been studied a great deal. All the multirate model-based inferential control schemes assumed that the fast-sampled process model was available, but unfortunately, no method had been proposed to estimate the fast-sampled process model from multirate data which is often the only source to obtain the multirate process model. One of the purposes of this thesis is to fill the gap by providing methods for estimating the fast sampled single-rate process model from multirate input-output data.

In most of the industrial applications, MPC schemes have been applied to multi-input and multi-output (MIMO) systems. It is well known that it is very efficient to represent a MIMO system in the state space framework. Subspace based system identification involves estimating the state space model of a process from the input-output data directly. The proposed multirate identification schemes were first developed in subspace framework then discussed in the polynomial domain. The unique problems in multirate identification such as the accuracy of time-delay estimation, the observability in the presence of time-delay and the causality problem were discussed in detail. Three methods to extract a fast sampled model from the lifted model were presented. The multirate identification schemes have been applied to both SISO and MIMO multirate systems. The multirate identification algorithms were analyzed in frequency domain; this results in a new and effective way to design excitation sequence for multirate identification. Experimental and simulation examples as well as industrial applications were used to demonstrate the multirate identification schemes.

In multirate systems, some of the output sampling interval may be so large that some loops can not be closed, but other loops, for economic or safety reasons, are closed at the same time. Traditional subspace-based schemes can not be applied to such systems which have both open-loop and closed-loop sub-systems. A multirate identification algorithm which can estimate the fast-sampled process model is developed for such systems where some loops are closed but the others are open. The proposed algorithm can compute the fast-sampled model in either one step or two steps. This identification scheme can handle the most general class of processes in contrast to that other subspace-based schemes can only be applied to the ARMAX (auto-regressive and moving-average, x denotes external input) type of processes.

A multirate inferential model-based predictive control scheme is proposed in both polynomial and state space domains. Model-plant mismatch (MPM) is inevitable in chemical industry due to various factors. Analysis of the robust stability and performance of the multirate inferential control systems in the presence of MPM is very important. The nominal performance and robust stability of a special class of mulitrate inferential control systems were theoretically analyzed. Experimental and simulation examples were explored to illustrate the performance and stability of the multirate inferential control schemes.

Acknowledgements

I would like to express my gratitude towards my supervisors, Prof. Sirish Shah and Prof. Tongwen Chen, for their constant encouragement, help and guidance. I am grateful to Prof. Shah for inspiring my interests in research and giving me the freedom in the choice of my research area. I will always cherish Prof. Shah's valuable advice, especially his emphasis on effective communication, and thankful for his additional financial support for my English courses. I am indebted to Prof. Chen for the invaluable time he has spent on correcting my reports, and giving me a few grammar lessons.

I would like to thank Prof. Huang for teaching me the basics of performance assessment and model-based predictive control. I am grateful to Prof. Forbes for giving me the opportunity to use the computing facilities in the initial stages of my doctoral program. I am thankful to all those who have taught me the fundamentals of the relevant areas of my research.

I wish to thank Dr. Lakshminarayanan for teaching me the concepts of CVA, and other statistical techniques. I would like to express my gratitude to Arun for helping me with the software, and the stimulating discussions we shared. I wish to thank Dr. Weihua Li for giving me valuable advice and correcting my report. I am thankful to Dr. Kent Qi of Shell Canada, Fort Sasketchwan for giving me the opportunity to apply my research ideas to industrial problems.

Through my doctoral program, I have been enjoying the friendly and exciting environment in the process control group. Biao, Laksh, Rohit, Anand, Aseema, Arun, Misha, Kamrun, Daniel, Bhushan, Remesh, Haitao, Yale, Sachin, Huilan, Xing, Weihua, Shoukat, Ashish, Jiangping, Lanny, Abhishek and Zhengang have all been good friends and great teachers to me. I have had great pleasure in sharing my office with Bhushan, and thoroughly enjoyed the interesting discussions on a broad range of topics.

I would like to gratefully acknowledge the department of chemical & materials engineering for giving me the opportunity to pursue my doctoral degree and providing an atmosphere most congenial for my research. Special thanks to the computing staff, Bob and Jack, for their timely help with computer related matters and the administrative staff for their warm cooperation and help.

I will always carry with me the memory of the good time I spent with my best friends, Xiaoming, Shuyin, Jianjun, Lan Li, Weishan, Jiang Bai and Bin Cai.

Finally, but most importantly, I am immensely grateful to my family members. The constant encouragement and affection from my parents has inspired me to this

extend and shall be remembered forever. Xuechao, my wife, has been a wonderful partner and a constant source of motivation, both during the good and bad times.

Contents

1 Introduction 1

1.1 Multirate Control . . . 1

1.2 Multirate Identification . . . 4

1.3 Robust Stability Analysis of Multirate Inferential Control Systems . . 5

1.4 Thesis Outline . . . 5

2 Subspace-based Identification Methods 8

2.1 Introduction . . . 8

2.2 Canonical Variate Analysis . . . 9

2.2.1 Canonical Correlation Analysis . . . 10

2.2.2 CVA Identification Algorithm . . . 11

2.3 N4SID . . . 14

2.4 MOESP . . . 20

2.5 Unifying Framework . . . 23

2.6 Illustrative Examples . . . 24

2.6.1 Experimental Example . . . 24

2.6.2 Simulation Example . . . 27

3 Introduction to Multirate Systems 33

3.1 The Lifting Technique and Lifted Systems . . . 33

3.2 Multirate Control . . . 37

3.2.1 Slow-Rate Control . . . 37

3.2.2 Lifted Control . . . 37

3.2.3 Inferential Control . . . 39

4 Subspace-based Identification of SISO Multirate systems 41

4.1 Introduction . . . 41

4.2 Lifted Systems . . . 42

4.3 Observability in the Presence of Time Delay . . . 46

4.4 Time-Delay Estimation 49
4.5 Identification of Lifted Models 51
4.6 Computing Fast Single-Rate Models 54
4.6.1 Controllability Approach 55
4.6.2 Eigenvalue Approach 55
4.6.3 Matrix Roots Approach 56
4.7 Procedure Summary 59
4.8 Simulation Example 60
4.9 Conclusion 61

5 Subspace-based Identification of MIMO Multirate Systems 62
5.1 Introduction 62
5.2 Identification of MIMO Multirate Systems 62
5.2.1 Lifted MIMO Multirate Systems 64
5.2.2 Extracting Fast Single-rate Models 68
5.3 Simulation and Experimental Examples 70
5.3.1 A Simulation Example 70
5.3.2 An Experimental Example 70
5.4 Conclusion 74

6 Multirate System Identification in the Polynomial Domain 75
6.1 Introduction 75
6.2 Multirate System Identification in the Transfer Function Domain . . . 76
6.3 Analysis in the Frequency Domain 79
6.4 Industrial Case Study 81
6.5 Conclusion 85

7 System Identification of Closed-loop Multirate Systems 87
7.1 Introduction 87
7.2 Preliminaries 89
7.3 Identification of Open-loop Single-rate Systems 91
7.3.1 Estimation of A_d and C_d 93
7.3.2 Estimation of B_d and D_d 95
7.4 Identification of Closed-loop Single-rate Systems 96
7.5 Identification of Closed-loop Multirate Systems 99
7.6 Noise Effect On Estimation 103
7.6.1 Open-loop Case 103
7.6.2 Closed-loop Case 104

7.7 Consistency Analysis . . . 105
7.7.1 Open-loop Analysis . . . 105
7.7.2 Closed-loop Analysis . . . 107
7.8 Illustrative Examples . . . 107
7.8.1 Experimental Examples . . . 107
7.8.2 Simulation Examples . . . 111
7.9 Conclusion . . . 113

8 Multirate Inferential MPC **116**
8.1 Introduction . . . 116
8.2 Inferential MPC . . . 118
8.2.1 Inferential Control Algorithm in the Transfer Function Domain 118
8.2.2 Inferential Control Algorithm in the State-Space Domain . . . 122
8.3 Experimental Evaluation . . . 124
8.4 Conclusion . . . 126

9 Analysis of Multirate Inferential Systems **127**
9.1 Introduction . . . 127
9.2 Preliminaries . . . 128
9.3 Dual-Rate Inferential Control Scheme . . . 129
9.4 Nominal Performance . . . 131
9.5 Stability Robustness . . . 135
9.6 Illustrative Example . . . 137
9.7 Conclusion . . . 139

10 Conclusions and Future Work **140**
10.1 Conclusions . . . 140
10.2 Future Work . . . 141

List of Figures

1.1 *A simplified block diagram of a multirate sampled-data system* 2

2.1 *Geometric interpretation of CCA* 11
2.2 *Block diagram of a deterministic-stochastic sampled-data system* . . . 12
2.3 *Block diagram of a stirred tank heater system* 25
2.4 *Block diagram for model validation* 26
2.5 *The measured water level and the output of the model estimated using the CVA method* 26
2.6 *The measured water temperature and the output of the model estimated using the CVA method* 26
2.7 *The measured water level and the output of the model estimated using the N4SID method* 27
2.8 *The measured water temperature and the output of the model estimated using the N4SID method* 27
2.9 *The measured water level and the output of the model estimated using the MOESP method* 28
2.10 *The measured water temperature and the output of the model estimated using the MOESP method* 28
2.11 *The step responses of the process from u_1 to y_1 and the model estimated using the CVA method* 29
2.12 *The step responses of the process from u_1 to y_2 and the model estimated by using the CVA method* 29
2.13 *The step responses of the process from u_2 to y_2 and the model estimated by using the CVA method* 30
2.14 *The step responses of the process from u_1 to y_1 and the model estimated by using the N4SID method* 30
2.15 *The step responses of the process from u_1 to y_2 and the model estimated by using the N4SID method* 30

2.16 *The step responses of the process from u_2 to y_2 and the model estimated by using the N4SID method* 31
2.17 *The step responses of the process from u_1 to y_1 and the model estimated by using the MOESP method* 31
2.18 *The step responses of the process from u_1 to y_2 and the model estimated by using the MOESP method* 31
2.19 *The step responses of the process from u_2 to y_2 and the model estimated by using the MOESP method* 32

3.1 *Block diagram of a SISO multirate sampled-data system with fast hold* 33
3.2 *Block diagram of a SISO lifted multirate sampled-data system* 34
3.3 *Block diagram of a SISO multirate sampled-data system with slow hold* 36
3.4 *Block diagram of a lifted SISO multirate sampled-data system with slow hold* 36
3.5 *Block diagram of a SISO multirate sampled-data system with a slow SR controller* 38
3.6 *Block diagram of a SISO multirate sampled-data system with a multirate controller* 38
3.7 *Block diagram of a SISO multirate sampled-data system with an inferential controller* 40

4.1 *Block diagram of a SISO multirate sampled-data system* 43
4.2 *The lifted SISO multirate system* 43
4.3 *Comparison between the true fast model and the fast model extracted by using the controllability approach* 48
4.4 *Comparison between the true fast model and the fast model extracted by using the matrix roots approach* 49
4.5 *Step responses of the actual (solid) and estimated (star) fast-rate models* 60

5.1 *Block diagram of a MIMO multirate sampled-data system* 63
5.2 *The lifted MIMO multirate system* 64
5.3 *Step responses of the true model (G11) and the estimated model* . . . 70
5.4 *Step responses of the true model (G12) and the estimated model* . . . 71
5.5 *Step responses of the true model (G21) and the estimated model* . . . 71
5.6 *Step responses of the true model (G22) and the estimated model* . . . 71
5.7 *Experimental setup of a stirred tank heater* 72
5.8 *Water level plots for the model (star) and the actual process (solid)* . 73
5.9 *Water temperature plots for the model (dot) and the actual process (solid)* 73

6.1 *Block diagram of a special SISO multirate sampled-data system* 77
6.2 *A special SISO multirate sampled-data system with the "vectorized" excitation* 81
6.3 *Output and manipulated variable measurements* 83
6.4 *Three disturbances* 84
6.5 *The output measurement (solid line) and the prediction of the fast model (dashdot line* 85
6.6 *Comparison of performance before and after implementing automatic control* 86

7.1 *Block diagram of a general combined deterministic-stochastic sampled-data system* 89
7.2 *The block diagram of a SR closed-loop sampled-data system* 97
7.3 *The block diagram of a closed-loop sampled-data system* 99
7.4 *The block diagram of a closed-loop MR sampled-data system* 100
7.5 *Block diagram of a stirred tank heater system* 108
7.6 *Measured and simulated water levels* 109
7.7 *Measured and simulated water temperatures* 109
7.8 *Block diagram of the stirred tank heater system with an inferential control scheme* 110
7.9 *Measured and simulated water level* 110
7.10 *Measured and simulated water temperature* 111
7.11 *Performance of the stirred tank heater with an MR inferential GPC* . 111
7.12 *Step response of the true system and the estimated system* 112
7.13 *Bode plot of the true system and the estimated system* 112
7.14 *The step responses of the process (relating the water level to the cold water valve) and the estimated model* 113
7.15 *The step responses of the process (relating the water temperature to the cold water valve) and the estimated model* 114
7.16 *The step responses of the process (relating the water temperature to the steam valve) and the estimated model* 114

8.1 *Block diagram of a SISO multirate sampled-data system with a lifted MPC* 117
8.2 *Block diagram of a SISO multirate sampled-data system with an inferential MPC* 117
8.3 *Experimental setup of a stirred tank heater* 124
8.4 *The measured water level* 125

8.5 *The measured water temperature* 125

9.1 *The sampled-data single-rate control system.* 129
9.2 *The discrete-time single-rate control system.* 129
9.3 *The sampled-data inferential control system.* 130
9.4 *The discrete-time inferential control system.* 131
9.5 *The lifted inferential control system.* 133
9.6 *The inferential control system with additive uncertainty.* 135
9.7 *The lifted inferential control system with additive uncertainty.* 136
9.8 *The equivalent system of Figure 9.7.* 136
9.9 *Step response in y for the fast single-rate system.* 138
9.10 *Step response in y for the dual-rate inferential control system.* 138

Chapter 1

Introduction

Multirate (MR) systems are sampled-data systems with non-uniform sampling/hold intervals. Figure 1.1 shows a simplified block diagram of multirate systems. Multirate systems are very common in the chemical process industry; typical examples include distillation columns and reactors. In distillation columns or reactors, the primary measurements such as the composition, density or molecular weight distribution measurements take several minutes or even several hours of analysis time; on the other hand, secondary measurements such as flow-rates, temperatures and pressures are sampled at relatively fast rate; and usually the manipulative variables can be adjusted at relatively fast sample rates. These naturally lead to multirate sampled-data systems.

Single-rate (SR) systems, whose signals are all sampled at one identical rate, can be considered as a special case of multirate systems. Single-rate systems are linear time-invariant (LTI) if the continuous systems are LTI; but multirate systems are time-varying regardless of the time-invariance of the continuous systems. Compared with single-rate systems, multirate systems are relatively more difficult in controller design and identification because the mathematical complexity involved in dealing with multirate systems is more intriguing than that involved in dealing with single-rate systems. This is the main reason why control and system identification techniques for multirate systems have evolved at a much slower pace.

1.1 Multirate Control

Study of multirate systems goes back to the early 1950's. The first important work was performed by Kranc in 1957 [43] on the switch decomposition technique; later, Friedland further developed the concept of lifting which converts a periodic discrete-time system into a time-invariant system [24]. Since then, the lifting technique has

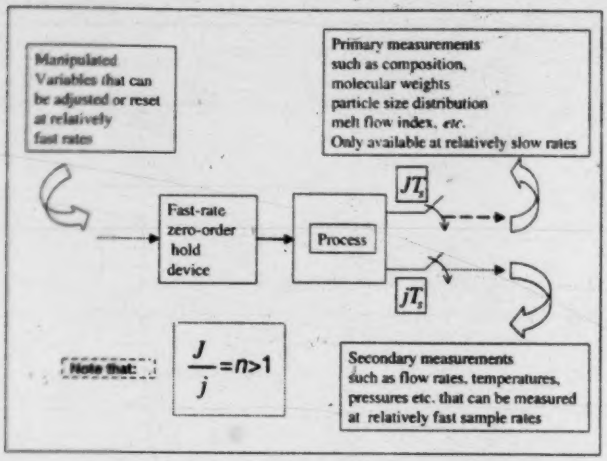

Manipulated Variables that can be adjusted or reset at relatively fast rates
Primary measurements such as composition, molecular weights particle size distribution melt flow index, etc. Only available at relatively slow rates
Fast-rate zero-order hold device
Process
JT_s
JT_s
Note that:
$\frac{J}{j}=n>1$
Secondary measurements such as flow rates, temperatures, pressures etc. that can be measured at relatively fast sample rates

Figure 1.1: *A simplified block diagram of a multirate sampled-data system*

become an important and widely used tool for analysis of multirate systems. The fundamental idea of lifting technique can be explain by a simple example: Consider a single-input single-output multirate system where the control interval is T and the output sampling interval is nT. The lifting technique takes the n control signals in time interval nT as n independent components of a vector control signal with interval nT, the resulting fictitious system which relates the vector input to the sampled output is single-rate and LTI but with increased input dimensionality. Clearly, by making use of the periodicity nature of the multirate systems, the lifting operator converts multirate systems to fictitious single-rate systems; such fictitious single-rate systems is the so-called the name lifted systems. Single-rate control theory can then be modified and applied to the lifted systems, and this gives rise to the *lifted control theory*. The lifting and inverse lifting operators provide a framework for controller design and analysis of multirate systems. Numerous optimal H_2 and H_∞ control algorithms have been developed for several classes of multirate systems [9, 11, 66], in the last decade.

Besides the lifted control schemes, inferential control scheme which estimates the inter-sample outputs and then feedback the estimated inter-sample outputs to the controller is another way to achieve fast control movements for multirate systems. Traditionally, the inter-sample estimation is based on the fast sampled secondary measurements. For example, in a distillation column, since the measurement of the composition in the top tray takes longer period of time, temperature of certain tray which is supposed to be strongly correlated with the composition is used as the

secondary measurement to infer the primary composition measurement. The traditional inferential schemes has been applied in industry with some success [48], but the traditional inferential schemes which ignore the infrequently measured primary measurements often give biased estimates of the primary outputs in the presence of disturbances. Model-based inferential schemes which incorporate the infrequent primary output measurements into estimation schemes were first presented by Dhulster et al. [17]. Lu and Fisher studied the inter-sample outputs estimation and inferential control for such multirate systems where the ratio between the uniform output sampling interval and the uniform control interval is an integer [57, 58]; their estimation method is not based on the process model at interval T but the relationship between the future fast sampled outputs and past input-output measurements; they also showed the convergence properties of their method. Guilandoust et al. [29, 30] considered more general multirate models with slow sampled main outputs and showed that the inter-sample outputs can be estimated with the aid of a fast sampled secondary output. Lee and Morari developed a generalized inferential control scheme and discussed various optimal control problems in the LQG, MPC, and IMC framework [47]. Gudi and Shah have developed an enhanced observability estimation method for multirate processes [27]. Oisiovici and Cruz have developed a model-based inferential controller and applied it to a batch distillation column; the model they used is a linearized first-principle model [64]. The three methods developed by Guilandoust et al, Lee and Morari, Gudi and Shah assume that a fast sampled model of the process is available. This in itself can be a nontrivial task, as it requires estimation of the fast-sampled process dynamics from multirate data; one of the purposes of our work is to fill the gap by providing methods of obtaining such fast single-rate models based on multirate data.

In the past two decades process control practice has progressed from simple PID controllers to advanced model-based predictive controllers (MPC) [16, 15, 67] capable of handling interactions and constraints. The basic idea in the strategy of MPC is to minimize the sum of squares of future control errors over a finite-time horizon and at the same time pay attention to the incremental control energy required to minimize the control errors. In a receding horizon fashion, model based predictive controllers only implement the first of the calculated control moves. The ability to deal with interactions and constraints makes MPC widely accepted by the process industry.

Multirate systems where the uniform output sampling interval is nT and the uniform control interval is T are one of the simplest classes of multirate systems, but it is non-trivial to implement model-based predictive controller with a control interval of T even for such simple multirate systems. Clearly, if the single-rate fast-

sampled process model is available, then we can easily design a single-rate MPC with interval T which processes both the outputs measured every nT and the estimated inter-sample outputs; this is the basic idea of multirate inferential MPC. Dynamic Matrix Control (DMC) and Generalized Predictive Control (GPC) are two important and widely used classes of MPC applied in chemical industry. DMC and GPC type multirate inferential MPC will be developed in Chapter 8.

1.2 Multirate Identification

Theoretically process models can be derived from basic physical laws. But in practice, it is very often that the input-output measurement is the only source to obtain process model. System identification [55] is concerned with identifying the model of a process from input-output data. System identification in both the polynomial domain [55, 68] and the state-space domain [45, 75, 79] has achieved much success. But these identification algorithms can be applied only to single-rate input-output data.

Traditionally, engineers interpolate the inter-sample outputs from the slowly sampled outputs, then estimate fast single-rate models based on both the measured and interpolated outputs [1]. The accuracy of the fast models estimated by the traditional approach mainly relies on the accuracy of inter-sample interpolation which does not take process dynamics into account. Clearly, the more reasonable approach is to estimate the fast model directly from the multirate input-output data; we use the second approach in this thesis.

The estimation of fast single-rate models directly from multirate input-output data is a relatively new research area. Verhaegen and Yu presented a technique to estimate lifted models of multirate systems in the state-space domain [82]. In their work, they represented multirate systems with periodic systems, and they estimated the lifted model with the multi-variable output-error state-space method; their method can not handle the causality constraint which is crucial for identifying the lifted models. In our work, we handle the lifted systems which are linear time-invariant, so we can apply most of the existing identification algorithms to estimate the lifted systems with a little modification to overcome the causality constraint; and we go one step further to extract fast sampled models from the lifted models. Identification of fast sampled state-space models of the single-input and single-output (SISO) and multi-input and multi-output (MIMO) multirate systems will be discussed in Chapters 4 and 5, respectively. Multirate identification in the transfer function domain will be discussed in Chapter 6.

For safety reasons in industry, it is quite often that modeling and identification

can only be practiced under closed-loop conditions. The difficulty for closed-loop identification is that the control signals are correlated with the disturbances under closed-loop conditions which violates the assumption of most identification algorithms. Gustavsson and co-workers have surveyed some problems such as identifiability and accuracy etc. in closed-loop identification in the polynomial domain [31]. Forssell and Ljung have studied the statistical properties of the closed-loop identification in the prediction error framework [22]. Verhaegen has applied the multi-variable output-error state-space algorithm to closed-loop identification [78], but he identified the augmented system instead of the process model directly. All existing subspace identification algorithms have been developed for processes which can be represented by auto-regressive and moving-average (ARMA) models. A subspace-based closed-loop identification scheme for the most general class of processes [55] will be discussed in Chapter 7, and the proposed closed-loop identification algorithm is capable of processing multirate data.

1.3 Robust Stability Analysis of Multirate Inferential Control Systems

Clearly model-plant mismatch is inevitable; the designed controllers generally work well for the nominal processes. However when the controllers are implemented on real processes which deviate from the nominal operating conditions, the performance may become much worse than the desired performance, or the closed-loop systems may even become unstable. It is important to make sure that the closed-loop systems are stable when the controllers are implemented on all possible process operating conditions and that the performance is within a user specified range. This type of analysis and design method is called robust stability and performance analysis [61]. Robust stability and performance analysis of single-rate systems has been studied extensively, but the robust stability and performance analysis of multirate systems with inferential controllers has not been studied in detail. Theoretical analysis and some interesting results will be presented in Chapter 9.

1.4 Thesis Outline

In this thesis, we will focus on identification of multirate systems, multirate inferential control algorithms and robust stability and performance analysis of multirate inferential control systems. The thesis is organized as follows:

First we will give a tutorial introduction of three popular subspace based identification schemes in Chapter 2. All the three algorithms are illustrated with experimental and simulation examples.

Chapter 3 gives an introduction to the control of multirate systems. In this chapter, at first the lifting/inverse lifting operators and the lifted systems will be introduced, then the lifted control schemes will be discussed and compared with the inferential control schemes.

Chapter 4 proposes the fundamental idea of the two-step multirate system identification strategy: estimate the lifted model and then extract a fast sampled model from the lifted model. This chapter considers identification of the general class of SISO multirate processes where the ratio between the output sampling interval and control interval is a rational number. This chapter will focus on the observability of the lifted model in the presence of a large time-delay, causality constraint on the lifted model, and the methods for extracting the fast sampled model from the lifted model. The final result of Chapter 4 is that a single-rate state space model where the output is sampled as fast as the control interval can be estimated from the multirate data.

Chapter 5 extends the results in Chapter 4 to a special but practical class of MIMO multirate systems where all the control intervals are uniform, all the output sampling intervals are distinct, but all the output sampling intervals are integer multiples of the control interval. The most important result in Chapter 5 is that a single-rate state-space model of the process sampled with the control period can be estimated from the multirate data. Simulation and experimental examples are included to validate the results in Chapters 4 and 5.

Identification of multirate systems in the transfer function domain is elaborated in Chapter 6. This chapter consider only a special class of multirate systems where the output period is an integer multiple of the control period. First, two methods to estimate a fast sampled model from the multirate data are developed. Providing another dimension of insight into the processes, analysis of identification schemes in frequency domain is important and practical. Analysis of the proposed multirate identification is an interesting part of this Chapter. An industrial application of the proposed multirate identification method is presented at the end of this chapter.

In multirate systems, some of the output sampling intervals may be so large that some loops can not be closed, but other loops are closed at the same time. Traditional subspace-based schemes can not be applied to such open-/closed-loop systems. Multirate systems include both open loops and closed loops are not uncommon in chemical industry, so it is of great importance to develop subspace-based schemes

for identifying such systems. Chapter 7 gives a subspace-based instrumental variable method which can estimate the fast sampled model of the process under such open-/closed-loop conditions. One advantage of the proposed subspace-based identification method is that it can handle the most general class of processes. Another advantage is that it can handle multirate data, specifically, it can estimate the fast-sampled process model in either one step or two steps. Simulation and experimental examples are used to illustrate the proposed algorithm.

Theoretical analysis along with experimental work in Chapter 8 shows that the inferential control algorithm can be incorporated into the existing single-rate MPC such as DMC and GPC in both the state-space framework and the polynomial framework. Experimental examples are included to demonstrate the performance of the multirate inferential MPC.

In Chapter 9, the performance and stability of the inferential control scheme in the absence of model-plant mismatch is first discussed. The focus of this chapter, stability robustness of inferential control systems in the presence of model-plant mismatch, is then elaborated. An illustrative example is included to validate the results given.

Chapter 10 gives an overall picture of the work in this thesis, and proposes some future work.

Chapter 2

Subspace-based Identification Methods

2.1 Introduction

Control engineers in process industries have strong desire to obtain better knowledge of specific plants in order to obtain improved control and thus lower operating cost, since it is well known that the performance of the advanced control schemes such as Smith predictor, H_2-norm control, dynamic matrix control (DMC) and generalized predictive control (GPC), etc., relies on the accuracy of the models they based. Theoretically the dynamic process models can be derived from basic physical/chemical laws, but in practice, it is very often that the input and output measurement is the only source to obtain process models.

System identification [55] is concerned with identifying the model of a process from input-output data. Estimation of the transfer function representation of a process has achieve tremendous success for a long time, especially after being applied to estimate dynamic process models for the advanced model-based predictive control schemes. In most of the industrial applications, model-based predictive control schemes have been applied to MIMO systems. It is well known that it is very efficient to represent and analyze MIMO systems in the state-space framework. Subspace based system identification involves estimating the state-space models of processes from the input-output data directly.

Subspace-based system identification is a relatively new and active research area, and it has been very successful in industry after its advantage being understood. The subspace-based identification methods originate from the classical realization theory formulated by Ho and Kalman [34], and Kung [44]. A number of subspace identification algorithms have been published. Larimore presented the CVA (canonical variate

analysis) method in 1990 [45], which is based on maximum likelihood estimation. Numerical algorithms for subspace state space system identification (N4SID) were developed by Van Overschee and De Moor [75, 76], the key idea of N4SID being orthogonal projection. MOESP (multi-variable output-error state-space) methods were presented and discussed by Verhaegen in a series of his papers [79, 80]; MOESP is basically an instrumental variable method.

All the subspace identification algorithms share the same basic idea: extract the information of the state from the space spanned by the measured inputs and outputs. Specifically the subspace methods can be classified into two groups: one group estimates the state from the space spanned by the inputs and outputs first, then estimates the state-space model (A, B, C, D) by solving a least square problem [45, 76]; the other group estimates the extended observability matrix from the space spanned by the inputs and outputs, then compute the state-space model (A, B, C, D) [79]. Van Overschee and De Moor have explored the similarities between the three algorithms (CVA, N4SID and MOESP), and have shown that all three methods are special cases of one unifying theorem [77].

Compared with identification algorithms in the polynomial framework, subspace-based identification methods are simple and numerically more stable; and subspace identification algorithms for SISO systems can be extended to MIMO systems naturally. Another major advantage is that subspace-based identification algorithms are non-iterative, with no non-linear optimization part involved [77].

In this chapter we will first briefly introduce CVA, N4SID and MOESP methods, then use both simulation and experimental examples to illustrate the three algorithms.

2.2 Canonical Variate Analysis

The Canonical variate analysis (CVA) method, presented in 1990 by Larimore [45], is the first subspace-based identification algorithm. The key idea of the CVA identification method is to estimate the state from the space spanned by the inputs and outputs through canonical correlation analysis (CCA). Canonical correlation analysis was first developed and applied in linear algebra by Jordan in 1875, and later was introduced into the statistical community by Hotelling in 1936 [35]. Before discussing of the CVA identification method, we will first briefly introduce the canonical correlation analysis.

2.2.1 Canonical Correlation Analysis

Assume that there are two sets of random variables. The first set, which has m variables, is represented by the $(m \times 1)$ random vector X_1. The second set, which has l variables, is represented by the $(l \times 1)$ random vector X_2. We also assume that $m \leq l$. For simplicity we assume that all the variables have zero mean. Define

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$\Sigma = E(XX^T) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where E represents the statistical average. Define a linear combination in X_1 and a linear combination in X_2 as

$$U = a'X_1, \qquad V = b'X_2,$$

where a and b are vectors with appropriate dimensions. The objective is to seek a and b such that

$$Corr(U, V) = \frac{a'\Sigma_{12}b}{\sqrt{a'\Sigma_{11}a}\sqrt{b'\Sigma_{22}b}}$$

is as large as possible. Define

$$\begin{aligned} W_1 &\triangleq \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}, \\ W_2 &\triangleq \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}. \end{aligned}$$

Assume that $\gamma_1^2 \geq \gamma_2^2 \geq \cdots \geq \gamma_m^2$ are the eigenvalues of W_1, and e_i is the eigenvector associated with γ_i^2. It can be shown that γ_1^2, γ_2^2, $\cdots$, γ_m^2 are also the m largest eigenvalues of the matrix $\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$ with corresponding eigenvectors f_1, f_1, $\cdots$, f_m. Define

$$\begin{aligned} a_i' &= e_i'\Sigma_{11}^{-\frac{1}{2}}, \; b_i' = f_i'\Sigma_{22}^{-\frac{1}{2}}, \\ U_i &= a_i'X_1, \quad V_i = b_i'X_2. \end{aligned}$$

Then

$$\gamma_i = Corr(U_i, V_i).$$

U_i and V_i are the so called canonical variate, and have the following properties:

$$Var(U_i, U_i) = 1, \quad Var(V_i, V_i) = 1$$

$$Cov(U_i, U_j) = 0, \quad i \neq j$$

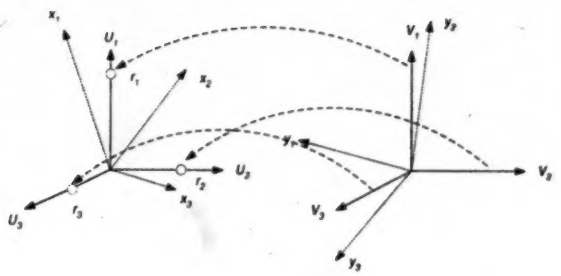


Figure 2.1: *Geometric interpretation of CCA*

$$Cov(V_i, V_j) = 0, \quad i \neq j$$

$$Cov(U_i, V_j) = 0, \quad i \neq j$$

for $i, j = 1, 2, ..., m$. γ_1 is the largest $Corr(U, V)$. The detail of the proof can be found in [38]. Assume that $m = l = 3$, then we can give some geometric interpretation of the canonical correlation analysis (CCA): Consider the projection as covariance, Figure 2.1 is the geometric interpretation of CCA. In Figure 2.1, x_i, i=1,2,3, represent the *ith* component of X_1; y_i, i=1,2,3, represent the *ith* component of X_2; original variables in X_1 and X_2 are dashed arrows; canonical variables are represented by solid lines. The canonical variables are uncorrelated, so they are orthogonal to each other; on the other hand, the original variables are correlated so they are not orthogonal.

2.2.2 CVA Identification Algorithm

The CVA method has the following assumptions:

- The observations are equal spaced in time.
- The system is finite-dimensional, linear and time-invariant.
- The noise disturbances are finite dimensional Gaussian processes.

Consider the single-rate combined deterministic-stochastic sampled-data system in Figure 2.2. In the block diagram, P_c is a continuous-time process with additive noise; the noise is generated by a continuous-time model N with a white noise input e; the output of P_c is corrupted by that of N, and is sampled by a sampler S_T with period T, yielding the sampled output $y(k)$; the input to P_c is generated by a zero-order hold with period T processing the input sequence $u(k)$. Both $u(k)$ and $y(k)$

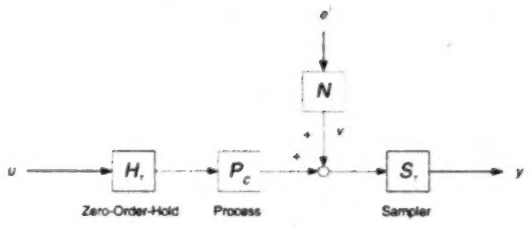


Figure 2.2: *Block diagram of a deterministic-stochastic sampled-data system*

are discrete-time signals defined on the time set $\mathcal{Z}_+ := \{0, 1.2, \cdots\}$ with underlying period T. Assume that the discrete system in Figure 2.2 can be represented by the following state space model

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + w(k) \\ y(k) &= Cx(k) + Du(k) + Fw(k) + v(k), \end{aligned}$$

where $x(k)$ is a n^{th} order Markov state, $w(k)$ and $v(k)$ are white noise sequences that are independent with covariance matrices Q and R, respectively; We have $A \in \mathbf{R}^{n\times n}$, $Q \in \mathbf{R}^{n\times n}$, $B \in \mathbf{R}^{n\times m}$, $C \in \mathbf{R}^{l\times m}$, $u \in \mathbf{R}^{m\times 1}$, $y, v \in \mathbf{R}^{l\times 1}$ and $w \in \mathbf{R}^{n\times 1}$. The objective is to estimate the coefficient matrices A, B, C, D and F along with the covariance matrices Q and R from the observed inputs $u(k)$ and outputs $y(k)$. Clearly if the state x can be estimated, then it is relatively easy to compute both the coefficient matrices and the covariance matrices.

We know that the basic idea of the CVA approach is to extract the information of the state x from the space spanned by the outputs and inputs. The estimate of the state involves a fundamental concept in the CVA approach and other subspace-based methods: the *past* and *future* of a process. At each time instant kT, a past vector p_k consisting of past outputs and inputs occurring prior to time kT and a future vector f_k consisting of outputs and inputs at time kT or later can be defined as follows

$$p_k = \begin{bmatrix} y_{k-1} \\ \vdots \\ y_{k-\alpha} \\ u_{k-1} \\ \vdots \\ u_{k-\alpha} \end{bmatrix}, \quad f_k = \begin{bmatrix} y_k \\ y_{k+1} \\ y_{k+2} \\ \vdots \\ y_{k+\alpha-1} \end{bmatrix},$$

where $\alpha (> n)$ is an integer. Assume that

$$\Sigma = E\left\{ \begin{bmatrix} f_k \\ p_k \end{bmatrix} \begin{bmatrix} f_k^T & p_k^T \end{bmatrix} \right\} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

The canonical correlation matrix between the future and past is then

$$\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}.$$

Singular value decomposition (SVD) of the canonical correlation matrix yields

$$\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}} = \begin{pmatrix} U_1 & U_2 \end{pmatrix}\begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix}\begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix},$$

where S_1 contains the n largest singular values, U_1 contains the first n columns of the left singular vector matrix, and V_1^T contains the first n rows of the right singular vector matrix. Larimore proved that

$$\hat{x}(k) = U_1^T p_k$$

is the maximum likelihood estimate of the true state $x(k)$ [45]. The input and output measurements together with the estimated state easily give the least square solution of the state-space coefficient matrices

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \sum\left[\begin{pmatrix} \hat{x}(k+1) \\ y(k) \end{pmatrix}, \begin{pmatrix} \hat{x}(k) \\ u(k) \end{pmatrix}\right]\left\{\sum\left[\begin{pmatrix} \hat{x}(k) \\ u(k) \end{pmatrix}, \begin{pmatrix} \hat{x}(k) \\ u(k) \end{pmatrix}\right]\right\}^{-1},$$

where

$$\begin{aligned} \sum\left[\begin{pmatrix} \hat{x}(k+1) \\ y(k) \end{pmatrix}, \begin{pmatrix} \hat{x}(k) \\ u(k) \end{pmatrix}\right] &= E\left[\begin{pmatrix} \hat{x}(i+1) \\ y(i) \end{pmatrix}\begin{pmatrix} \hat{x}(i) \\ u(i) \end{pmatrix}^T\right] \\ \sum\left[\begin{pmatrix} \hat{x}(k) \\ u(k) \end{pmatrix}, \begin{pmatrix} \hat{x}(k) \\ u(k) \end{pmatrix}\right] &= E\left[\begin{pmatrix} \hat{x}(i) \\ u(i) \end{pmatrix}\begin{pmatrix} \hat{x}(i) \\ u(i) \end{pmatrix}^T\right]. \end{aligned}$$

After estimating the coefficient matrices (A, B, C, D), it is straightforward to estimate the covariance matrices, (Q, R), and the coefficient matrix F:

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \sum\left[\begin{pmatrix} \hat{x}(k+1) \\ y(k) \end{pmatrix}, \begin{pmatrix} \hat{x}(k+1) \\ y(k) \end{pmatrix}\right] - \begin{bmatrix} A & B \\ C & D \end{bmatrix}\sum\left[\begin{pmatrix} \hat{x}(k+1) \\ y(k) \end{pmatrix}, \begin{pmatrix} \hat{x}(k) \\ u(k) \end{pmatrix}\right]$$

and the coefficient matrix F, and the covariance matrices Q and R are expressed as

$$\begin{aligned} F &= S_{21}S_{11}^{\dagger} \\ Q &= S_{11} \\ R &= S_{22} - S_{21}S_{11}^{\dagger}S_{12}, \end{aligned}$$

where $()^{\dagger}$ denotes the pseudoinverse.

2.3 N4SID

Van Overschee and De Moor developed the numerical algorithms for subspace state space identification (N4SID) in 1993 [76]. The N4SID approach is based on geometric projections, specifically, the N4SID first projects the future outputs onto the past outputs, past inputs and future inputs, then estimates the extended observability matrix from the projection (which is also a space spanned by the inputs and outputs); after that the N4SID estimates the state-space model. Compared with the CVA approach, the N4SID approach is mathematically more involved.

The N4SID approach also studies the combined deterministic-stochastic sampled-data system in Figure 2.2. The N4SID approach assumes that the discrete system in Figure 2.2 can be expressed in terms of the following state-space model

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + w(k) \\ y(k) &= Cx(k) + Du(k) + v(k), \end{aligned}$$

with

$$\begin{aligned} E\left[\begin{pmatrix} w(i) \\ v(i) \end{pmatrix}\begin{pmatrix} w(i) \\ v(i) \end{pmatrix}^T\right] &= \lim_{\beta\to\infty}\frac{1}{\beta}\sum_{i=1}^{\beta}\begin{pmatrix} w(i) \\ v(i) \end{pmatrix}\begin{pmatrix} w(i) \\ v(i) \end{pmatrix}^T \\ &= \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}. \end{aligned}$$

Let the input u, output y and state x have the same dimensions as in the CVA approach. We can see that the model assumed by the N4SID approach looks a little different from that assumed by the CVA approach, but in fact, the model used by the N4SID approach can easily be transformed into that used by the CVA approach through the Kalman filter approach [68]. The following assumptions are assumed to be satisfied:

- The observations are equal spaced in time.
- The system is finite-dimensional, linear and time-invariant.
- The noise disturbances are finite dimensional Gaussian processes.
- The pair (A, C) is observable, and $[A, (BQ^{1/2})]$ is controllable.
- The deterministic inputs are independent of the disturbances.

In order to explain the geometric projections clearly, let us first divide the deterministic-stochastic model into two subsystems: the deterministic subsystem,

$$\begin{aligned} x_d(k+1) &= Ax_d(k) + Bu(k) \\ y_d(k) &= Cx_d(k) + Du(k), \end{aligned}$$

and the stochastic subsystem,

$$\begin{aligned} x_s(k+1) &= Ax_s(k) + w(k) \\ y_s(k) &= Cx_s(k) + v(k). \end{aligned}$$

Define

$$\begin{aligned} U_{0|i-1} &\triangleq \begin{pmatrix} u(0) & u(1) & \cdots & u(J-1) \\ u(1) & u(2) & \cdots & u(J) \\ \vdots & \vdots & \cdots & \vdots \\ u(i-1) & u(i) & \cdots & u(i+J-2) \end{pmatrix}, \\ X_d(i) &\triangleq \begin{pmatrix} x_d(i) & x_d(i+1) & \cdots & x_d(i+J-1) \end{pmatrix}, \end{aligned}$$

$$\Gamma_i \triangleq \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{i-1} \end{pmatrix}, \; H_i \triangleq \begin{pmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \cdots & D \end{pmatrix},$$

and

$$\Delta_i = \begin{pmatrix} A^{i-1}B & A^{i-2}B & \cdots & B \end{pmatrix},$$

where $i(>n)$ is an integer, J is the observation number and Γ_i is the so called extended observability matrix. Define $Y_{0|i-1}$ and $Y^s_{0|i-1}$ by replacing u in $U_{0|i-1}$ with y and y_s, respectively. It is easy to obtain the relations:

$$\begin{aligned} Y_{0|i-1} &= \Gamma_i X_d(0) + H_i U_{0|i-1} + Y^s_{0|i-1}, \\ Y_{i|2i-1} &= \Gamma_i X_d(i) + H_i U_{i|2i-1} + Y^s_{i|2i-1}, \\ X_d(i) &= A^i X_d(0) + \Delta_i U_{0|i-1}. \end{aligned}$$

The N4SID approach also involves the concepts, *past* and *future*. For example, $U_{0|i-1}$ denotes the 'past' inputs and $U_{i|2i-1}$ denotes the 'future' inputs. The main project performed in the N4SID approach is defined as

$$\begin{aligned} Z_i &= Y_{i|2i-1} / \begin{pmatrix} U_{0|2i-1} \\ Y_{0|i-1} \end{pmatrix} \\ Z_{i+1} &= Y_{i+1|2i-1} / \begin{pmatrix} U_{0|2i-1} \\ Y_{0|i} \end{pmatrix}, \end{aligned}$$

where matrix division is defined as follows: for two matrices M_1 and M_2 of compatible dimensions,

$$M_1/M_2 \triangleq M_1 M_2^T (M_2 M_2^T)^{-1} M_2.$$

The main projection together with the assumption that the inputs u are independent of the noise/disturbance gives

$$Z_i = \Gamma_i \hat{X}(i) + H_i U_{i|2i-1} \tag{2.1}$$

$$Z_{i+1} = \Gamma_{i-1} \hat{X}(i+1) + H_{i-1} U_{i+1|2i-1} \tag{2.2}$$

with

$$\hat{X}(i) = \left(A^i - Q_i \Gamma_i \mid \Delta_i - Q_i H_i \mid Q_i \right) \left(\begin{array}{c} SR^{-1} U_{0|2i-1} \\ \hline U_{0|i-1} \\ \hline Y_{0|i-1} \end{array} \right) \tag{2.3}$$

$$\hat{X}(i+1) = \left(A^{i+1} - Q_{i+1} \Gamma_{i+1} \mid \Delta_{i+1} - Q_{i+1} H_{i+1} \mid Q_{i+1} \right) \left(\begin{array}{c} SR^{-1} U_{0|2i-1} \\ \hline U_{0|i} \\ \hline Y_{0|i} \end{array} \right) \tag{2.4}$$

and

$$\begin{aligned} Q_i &= \chi_i \psi_i^{-1}, \; P_d = E\left(x_d x_d^T\right) \\ P_s &= E\left(x_s x_s^T\right), \; G = A P_s C^T + S \\ \Delta_i^s &= \left(A^{i-1} G \;\; A^{i-2} G \;\; \cdots \;\; G \right) \\ \chi_i &= A^i (P_d - S R^{-1} S^T) \Gamma_i^T + \Delta_i^s \\ \psi_i &= \Gamma_i (P_d - S R^{-1} S^T) \Gamma_i^T + L_i^s. \end{aligned}$$

The proof of equations (2.3) and (2.4) can be found in [76]. Van Overschee and De Moor proved that the j columns of $\hat{X}_i$ are equal to the outputs of a bank of j non-steady state Kalman filters in parallel [76]. The proof will only be briefly illustrated here: Given $\hat{x}(0)$, P_0, $u(0)$, ..., $u(k-1)$, $y(0)$, ..., $y(k-1)$ and all the system matrices (A, B, C, D, Q, S, R), then the non-steady state Kalman filter state $\hat{x}(k)$ is defined by the following recursive equations,

$$\begin{aligned} \hat{x}(k) &= A\hat{x}(k-1) + Bu(k-1) + K_{k-1}[y(k-1) - C\hat{x}(k-1) - Du(k-1)] \\ K_{k-1} &= (AP_{k-1}C^T + G)(\Lambda_0 + CP_{k-1}C^T)^{-1} \\ P_k &= AP_{k-1}A^T - (AP_{k-1}C^T + G)(\Lambda_0 + CP_{k-1}C^T)^{-1}(AP_{k-1}C^T + G)^T, \end{aligned}$$

where

$$\Lambda_0 \triangleq E\left(y_s y_s^T\right).$$

The Kalman state can be also expressed as

$$\hat{x}(k) = \left(A^k - Q_k\Gamma_k \mid \Delta_k - Q_kH_k \mid Q_k \right) \left(\begin{array}{c} \hat{x}(0) \\ \hline u(0) \\ \vdots \\ u(k-1) \\ \hline y(0) \\ \vdots \\ y(k-1) \end{array} \right), \tag{2.5}$$

where

$$\begin{aligned} Q_k &= \chi_k\psi_k^{-1} \\ \chi_k &= A^iP_0\Gamma_k^T + \Delta_k^s, \end{aligned}$$

and

$$\psi_k = \Gamma_iP_0\Gamma_k^T + L_k^s.$$

Compare equations (2.3), (2.4) and (2.5), it is not difficult to conclude that the j columns of $\hat{X}_i$ are equal to the outputs of a bank of j non-steady state Kalman filters in parallel. The detail of the discussion is in [76]. From equations (2.1), (2.3) and (2.4) we can get

$$Z_i = \left(L_i^1 \mid L_i^2 \mid L_i^3 \right) \left(\begin{array}{c} U_{0|i-1} \\ \hline U_{i|2i-1} \\ \hline Y_{0|i-1} \end{array} \right) \tag{2.6}$$

$$Z_{i+1} = \left(L_{i+1}^1 \mid L_{i+1}^2 \mid L_{i+1}^3 \right) \left(\begin{array}{c} U_{0|i} \\ \hline U_{i+1|2i-1} \\ \hline Y_{0|i} \end{array} \right) \tag{2.7}$$

with

$$\begin{aligned} L_i^1 &= \Gamma_i(A^i - Q_i\Gamma_i)S(R^{-1})_{1|mi} + \Delta_i - Q_iH_i \\ L_i^2 &= H_i + \Gamma_i(A^i - Q_i\Gamma_i)S(R^{-1})_{mi+1|2mi} \\ L_i^3 &= \Gamma_iQ_i. \end{aligned}$$

Now it is evident that the following matrices have the same column space as Γ_i:

$$L_i^1 + L_i^3L_i^2 \tag{2.8}$$

and

$$\left(L_i^1 \mid L_i^3 \right) \left(\begin{array}{c} U_{0|i-1} \\ \hline Y_{0|i-1} \end{array} \right). \tag{2.9}$$

Singular value decomposition of either (2.8) or (2.9) can decide the model order n, and the extended observability matrix Γ_i. From (2.1) and (2.2) it follows that

$$\hat{X}(i) = \Gamma_i^{\dagger}[Z_i - H_i U_{i|2i-1}] \tag{2.10}$$

$$\hat{X}(i+1) = \Gamma_{i-1}^{\dagger}[Z_{i+1} - H_{i-1} U_{i+1|2i-1}], \tag{2.11}$$

and

$$\begin{pmatrix} \hat{X}(i+1) \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A \\ C \end{pmatrix} \hat{X}(i) + \begin{pmatrix} B \\ D \end{pmatrix} U_{i|i} + \begin{pmatrix} U_{0|2i-1} \\ Z_i \\ \hat{X}(i) \end{pmatrix}^{\perp}, \tag{2.12}$$

where $(.)^{\perp}$ indicates a matrix whose row space is perpendicular to the row space of (.). Substituting (2.10) and (2.11) into (2.12) gives

$$\begin{pmatrix} \Gamma_{i-1}^{\dagger} Z_{i+1} \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A \\ C \end{pmatrix} \Gamma_i^{\dagger} Z_i + \begin{pmatrix} \Re_{12} \\ \Re_{22} \end{pmatrix} U_{i|2i-1} + \begin{pmatrix} U_{0|2i-1} \\ Z_i \\ \hat{X}(i) \end{pmatrix}^{\perp} \tag{2.13}$$

where

$$\begin{pmatrix} \Re_{12} \\ \Re_{22} \end{pmatrix} = \begin{pmatrix} B - A\Gamma_i^{\dagger} \begin{pmatrix} D \\ \Gamma_{i-1} B \end{pmatrix} \Gamma_{i-1}^{\dagger} H_{i-1} - A\Gamma_i^{\dagger} \begin{pmatrix} 0 \\ H_i \end{pmatrix} \\ D - C\Gamma_i^{\dagger} \begin{pmatrix} D \\ \Gamma_{i-1} B \end{pmatrix} - C\Gamma_i^{\dagger} \begin{pmatrix} 0 \\ H_{i-1} \end{pmatrix} \end{pmatrix}.$$

After estimating n, Γ_i and Γ_{i-1} which is part of Γ_i, we can determine the least square solution, $(A, C, \Re_{12}, \Re_{22})$, of equation (2.13), and also compute matrices (B, D) from $(\Re_{12}, \Re_{22})$. This method is called N4SID method (1). There is also an approximate solution which is called N4SID method (2). The difference between N4SID methods (1) and (2) is that in method (2) the state sequences are calculated by dropping the linear combinations of $U_{i|2i-1}$ out of Z_i, and the linear combinations of $U_{i+1|2i-1}$ out of Z_{i+1}. In this way, the Kalman filter states of a different Kalman filter can be obtained. The resulting matrices are called $\Gamma_i \tilde{X}(i)$ and $\Gamma_{i-1} \tilde{X}(i+1)$ where

$$\Gamma_i \tilde{X}(i) = Z_i - L_i^2 U_{i|2i-1},$$

and

$$\Gamma_{i-1} \tilde{X}(i+1) = Z_{i+1} - L_{i+1}^2 U_{i+1|2i-1}.$$

Equations (2.6) and (2.7) yield

$$\tilde{X}_i = \left[\; (A^i - Q_i \Gamma_i) S (R^{-1})_{1|mi} + \Delta_i - Q_i H_i \mid Q_i \; \right] \begin{pmatrix} U_{0|i-1} \\ Y_{0|i-1} \end{pmatrix}$$

$$\tilde{X}_{i+1} = \left[\; (A^{i+1} - Q_{i+1} \Gamma_{i+1}) S (R^{-1})_{1|m(i+1)} + \Delta_{i+1} - Q_{i+1} H_{i+1} \mid Q_{i+1} \; \right] \begin{pmatrix} U_{0|i} \\ Y_{0|i} \end{pmatrix}.$$

If one of the following conditions,

- $i \to \infty$,
- u is a white noise,
- the system is purely deterministic,

is satisfied, then

$$\begin{aligned} \tilde{x}(k) &= \hat{x}(k) \\ \tilde{x}(k+1) &= \hat{x}(k+1). \end{aligned}$$

The procedure of method (2) can be summarized in the following steps:

- Calculate the projection

$$\begin{aligned} Z_i &= Y_{i|2i-1} / \begin{pmatrix} U_{0|i-1} \\ U_{i|2i-1} \\ Y_{0|i-1} \end{pmatrix} \\ &= \left(L_i^1 \mid L_i^2 \mid L_i^3 \right) \left(\begin{array}{c} U_{0|i-1} \\ \hline U_{i|2i-1} \\ \hline Y_{0|i-1} \end{array} \right) \end{aligned}$$

and

$$\begin{aligned} Z_{i+1} &= Y_{i+1|2i-1} / \begin{pmatrix} U_{0|i} \\ U_{i+1|2i-1} \\ Y_{0|i} \end{pmatrix} \\ &= \left(L_{i+1}^1 \mid L_{i+1}^2 \mid L_{i+1}^3 \right) \left(\begin{array}{c} U_{0|i} \\ \hline U_{i+1|2i-1} \\ \hline Y_{0|i} \end{array} \right). \end{aligned}$$

- Compute the singular value decomposition

$$\left(L_i^1 \mid L_i^3 \right) \left(\frac{U_{0|i-1}}{Y_{0|i-1}} \right) = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & \\ & S_2 \end{pmatrix} \begin{pmatrix} (V_1)^T \\ (V_2)^T \end{pmatrix}.$$

The model order n can be decided from the SVD (ideally n should be equal to the number of non-zero singular values, but due to disturbances, S_2 is not zero and we have to choose the model order using certain criteria). The extended observability matrix is

$$\Gamma_i = U_1 S_1^{1/2}.$$

Γ_{i-1} is simply part of Γ_i.

- Estimate the states $\tilde{X}_i$ and $\tilde{X}_{i+1}$ by calculating

$$\tilde{X}_i = \Gamma_i^{\dagger} \left(L_i^1 \mid L_i^3 \right) \begin{pmatrix} U_{0|i-1} \\ Y_{0|i-1} \end{pmatrix}$$

and

$$\tilde{X}_{i+1} = \Gamma_{i-1}^{\dagger} \left(L_{i+1}^1 \mid L_{i+1}^3 \right) \begin{pmatrix} U_{0|i} \\ Y_{0|i} \end{pmatrix}.$$

- Solve the least square problem,

$$\begin{pmatrix} \tilde{X}(i+1) \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A \\ C \end{pmatrix} \tilde{X}(i) + \begin{pmatrix} B \\ D \end{pmatrix} U_{i|i} + \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix},$$

and then

$$\begin{aligned} Q &= \rho_1 \rho_1^T \\ S &= \rho_1 \rho_2^T \\ R &= \rho_2 \rho_2^T. \end{aligned}$$

2.4 MOESP

The multi-variable output-error state-space identification algorithm (MOESP) was developed by Verhaegen in 1992 [80]. Similar to the N4SID approach, the MOESP also first estimates the extended observability matrix from the input-output space, then computes the state-space model. Compared with the N4SID, the MOESP is mathematically simpler and easier to understand. Consider the combined deterministic-stochastic sampled-data system in Figure 2.2, the MOESP assumes that the discrete combined system can be represented by such a state space model,

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Fw(k) \\ y(k) &= Cx(k) + Du(k) + Gw(k) + v(k), \end{aligned}$$

with

$$E\left[\begin{pmatrix} w(i) \\ v(i) \end{pmatrix} \begin{pmatrix} w(j) \\ v(j) \end{pmatrix}^T \right] = \begin{cases} \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

Here, u, y and x have the same dimensions as in the CVA approach. The MOESP makes the following assumptions:

- The observations are equal spaced in time.
- The system is finite-dimensional, linear and time-invariant.

- The noise disturbances are finite dimensional Gaussian processes.
- The pair (A, C) is observable, and $[A, (BQ^{1/2})]$ is controllable.
- The deterministic inputs are independent of the disturbances.
- The process is asymptotically stable.

Define

$$E_i \triangleq \begin{pmatrix} G & 0 & 0 & \cdots & 0 \\ CF & G & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ CA^{i-2}F & CA^{i-3}F & CA^{i-4}F & \cdots & G \end{pmatrix}$$

and

$$X(i) \triangleq \begin{pmatrix} x(i) & x(i+1) & \cdots & x(i+J-1) \end{pmatrix},$$

where $i(> n)$ is an integer and J is the observation number. It is straightforward to get the following equation

$$Y_{0|i-1} = \Gamma_i X(0) + H_i U_{0|i-1} + E_i W_{0|i-1} + V_{0|i-1}, \tag{2.14}$$

where $Y_{0|i-1}$ and $U_{0|i-1}$ have been defined in the discussion of the N4SID approach. Here $W_{0|i-1}$ and $V_{0|i-1}$ can be obtained by replacing u in $U_{0|i-1}$ with w and v, respectively. Applying QR decomposition to the input-output observations, we have

$$\begin{pmatrix} U_{i|2i-1} \\ U_{0|i-1} \\ Y_{0|i-1} \\ Y_{i|2i-1} \end{pmatrix} = \begin{pmatrix} R_{11} & 0 & 0 & 0 \\ R_{21} & R_{22} & 0 & 0 \\ R_{31} & R_{32} & R_{33} & 0 \\ R_{41} & R_{42} & R_{43} & R_{44} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \\ Q_4^T \end{pmatrix}. \tag{2.15}$$

Substituting (2.15) into (2.14) and its companion equation

$$Y_{i|2i-1} = \Gamma_i X(i) + H_i U_{i|2i-1} + E_i W_{i|2i-1} + V_{i|2i-1} \tag{2.16}$$

yields

$$\begin{aligned} \lim_{J\to\infty} \frac{1}{\sqrt{J}} Y_{i|2i-1} Q_2 &= R_{42} \\ &= \lim_{J\to\infty} \frac{1}{\sqrt{J}} \Gamma_i X(i) Q_2 \end{aligned} \tag{2.17}$$

and

$$\begin{aligned} \lim_{J\to\infty} \frac{1}{\sqrt{J}} Y_{i|2i-1} Q_3 &= R_{43} \\ &= \lim_{J\to\infty} \frac{1}{\sqrt{J}} \Gamma_i X(i) Q_3. \end{aligned} \tag{2.18}$$

From (2.17) and (2.18) it follows that the extended observability matrix, Γ_i, and the matrix $\begin{pmatrix} R_{42} & R_{43} \end{pmatrix}$ have the same column space [79]. So we can compute the extended observability matrix from the SVD of $\begin{pmatrix} R_{42} & R_{43} \end{pmatrix}$, specifically

$$\begin{pmatrix} R_{42} & R_{43} \end{pmatrix} = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & \\ & S_2 \end{pmatrix} \begin{pmatrix} (V_1)^T \\ (V_2)^T \end{pmatrix},$$

and then

$$\hat{\Gamma}_i = U_1 (S_1)^{1/2}.$$

After $\hat{\Gamma}_i$ being calculated, we can take the first l rows of $\hat{\Gamma}_i$ as matrix C and then calculate matrix A by solving such a least square problem:

$$\hat{\Gamma}_i(1:nl,:)A = \hat{\Gamma}_i(l+1:(n+1)l,:), \tag{2.19}$$

where n is the model order. The solution of (2.19) is clearly

$$A = [\hat{\Gamma}_i^T(1:nl,:)\hat{\Gamma}_i(1:nl,:)]^{-1}\hat{\Gamma}_i(1:nl,:)\hat{\Gamma}_i(l+1:(n+1)l,:). \tag{2.20}$$

So if the pair (C, A) is observable, then the matrix A computed from (2.20) is unique. From (2.14), (2.16) and (2.15) it also follows that

$$\begin{aligned} \lim_{J\to\infty} \frac{1}{\sqrt{J}} Y_{0|i-1} Q_1^T &= R_{31} \qquad (2.21) \\ &= \lim_{J\to\infty} \frac{1}{\sqrt{J}} (\Gamma_i X(0) Q_1^T + H_i R_{21}), \end{aligned}$$

$$\begin{aligned} \lim_{J\to\infty} \frac{1}{\sqrt{J}} Y_{0|i-1} Q_2^T &= R_{33} \qquad (2.22) \\ &= \lim_{J\to\infty} \frac{1}{\sqrt{J}} (\Gamma_i X(0) Q_2^T + H_i R_{22}), \end{aligned}$$

and

$$\begin{aligned} \lim_{J\to\infty} \frac{1}{\sqrt{J}} Y_{i|2i-1} Q_1^T &= R_{41} \qquad (2.23) \\ &= \lim_{J\to\infty} \frac{1}{\sqrt{J}} (\Gamma_i X(i) Q_1^T + H_i R_{11}). \end{aligned}$$

Combining (2.21), (2.22) and (2.23) gives

$$\begin{aligned} \lim_{J\to\infty} \frac{1}{\sqrt{J}} \begin{bmatrix} R_{31} & R_{32} & R_{41} \end{bmatrix} &= \lim_{J\to\infty} \frac{1}{\sqrt{J}} \{\Gamma_i \begin{bmatrix} X(0)Q_1^T & X(0)Q_2^T & X(i)Q_1^T \end{bmatrix} \qquad (2.24) \\ &\quad + H_i \begin{bmatrix} R_{21} & R_{22} & R_{11} \end{bmatrix}\}. \end{aligned}$$

Multiplying both sides of equation (2.24) by $\Gamma_i^\perp$ yields

$$\lim_{J\to\infty}\frac{1}{\sqrt{J}}\Gamma_i^\perp\left[\begin{array}{cccc} R_{31} & R_{32} & R_{41} \end{array}\right] = \lim_{J\to\infty}\frac{1}{\sqrt{J}}\Gamma_i^\perp H_i\left[\begin{array}{cccc} R_{21} & R_{22} & R_{11} \end{array}\right]. \tag{2.25}$$

If there is enough excitation then,

$$\Phi = \{\left[\begin{array}{cccc} R_{21} & R_{22} & R_{11} \end{array}\right]\left[\begin{array}{cccc} R_{21} & R_{22} & R_{11} \end{array}\right]^T\}$$

would be invertable, hence from (2.25) we can easily get

$$\lim_{J\to\infty}\frac{1}{\sqrt{J}}\Gamma_i^\perp\left[\begin{array}{cccc} R_{31} & R_{32} & R_{41} \end{array}\right]\left[\begin{array}{cccc} R_{21} & R_{22} & R_{11} \end{array}\right]^T\Phi^{-1} = \lim_{J\to\infty}\frac{1}{\sqrt{J}}\Gamma_i^\perp H_i. \tag{2.26}$$

Since $\Gamma_i^\perp$, C, A, and $\left[\begin{array}{cccc} R_{31} & R_{32} & R_{41} \end{array}\right]\left[\begin{array}{cccc} R_{21} & R_{22} & R_{11} \end{array}\right]^T\Phi^{-1}$ have been computed, so (2.26) is a linear equation of matrices B and D. It is straightforward to compute B and D from (2.26).

2.5 Unifying Framework

VanOverschee and DeMoor have explored the similarities between the N4SID, CVA and MOESP, and shown that all the three algorithms are special cases of one unifying theorem. Specifically, they proved that all of the three schemes used the same subspace to determine the order and the extended observability matrix, but used different weighting matrices to calculate the basis for the column space of the observability matrix. Their results in [77] are summarized in this section.

For an $(m \times n)$ real matrix $A = [a_{kl}]$, define

$$\|A\|_F^2 \triangleq \sum_{k=1}^{m}\sum_{l=1}^{n} a_{kl}^2,$$

the quantity $\|A\|_F$ is the so called Frobenius norm of A. The projection of the row space of matrix B onto the row space of matrix A is defined as

$$B/A \triangleq BA^T(AA^T)^{-1}A.$$

assuming the inverse exists. Let p represent the *past* inputs and outputs, u represent the *future* inputs and f represent the *future* outputs. All the three subspace-based identification algorithms solve the same optimization problems

$$\min \|W_1(\Phi - \Psi)W_2\|_F^2, \tag{2.27}$$

but with different weighting matrices W_1 and W_2. Φ in (2.27) is defined as follows:

$$\Phi = [(f/u^\perp)(p/u^\perp)^T][(p/u^\perp)(p/u^\perp)^T]^{-1}p.$$

With the singular value decomposition

$$W_1 \Phi W_2 = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & \\ & 0 \end{pmatrix} \begin{pmatrix} (V_1)^T \\ (V_2)^T \end{pmatrix},$$

the minimum-norm solution of of the optimization problem in (2.27) is

$$\Psi = W_1^{-1} U_1 S_1 V_1^T W_2^{\dagger}.$$

The extended observability matrix is equal to

$$\Gamma_i = W_1^{-1} U_1 S_1^{1/2}.$$

and the Kalman state sequence $\hat{x}$ can then be recovered from

$$\hat{x} = \Gamma_i^{\dagger} \Phi.$$

The weighting matrices of the three algorithms are summarized in the following table,

	N4SID	*MOESP*	*CVA*
W_1	I	I	$[(f/u^{\perp})(f/u^{\perp})^T]^{-1/2}$
W_2	I	$\Pi_{u^{\perp}}$	$\Pi_{u^{\perp}}$

Here $\Pi_{u^{\perp}}$ is

$$\Pi_{u^{\perp}} = (u^{\perp})^T (u^{\perp}(u^{\perp})^T)^{-1} u^{\perp}.$$

2.6 Illustrative Examples

A simulation and an experimental example will be presented in this section to illustrate the three subspace identification algorithms discussed in the former part of this chapter.

2.6.1 Experimental Example

The block diagram of a stirred tank heater process is shown in Figure 2.3. This process is a computer-controlled experimental setup at the University of Alberta. The steam through the steam pipe is used to heat the water in the tank; the cold water valve is used to adjust the water level in the tank; and the water outlet valve is fixed. An air-bubble stirrer is used to make the water in the tank well mixed. In this process, the two manipulated variables are the positions of the cold water (u_1) and the steam (u_2) valves; the two measured variables are the water level (y_1) and water temperature (y_2) in the tank. In this experiment, we use signal unit in milliampere

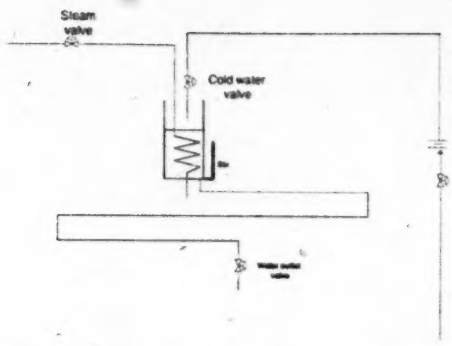


Figure 2.3: *Block diagram of a stirred tank heater system*

(mA) to quantify y_1, y_2, u_1 and u_2. The inputs and outputs of the stirred tank heater are summarized below:

Symbol	Quantity	Range
u_1	cold water valve	4 mA $\leq u_1 \leq$ 20 mA
u_2	steam valve	4 mA $\leq u_2 \leq$ 20 mA
y_1	water level	4 mA $\leq y_1 \leq$ 20 mA
y_2	water temperature	4 mA $\leq y_2 \leq$ 20 mA

A single-rate open-loop experiment was performed. In this experiment, all the valve positions were updated every 4 seconds, all the outputs were sampled every 4 seconds. An random binary vector sequence with 6000 data points was generated as excitation sequence. The normal output values were (14mA, 9mA). A total of 6000 experimental data points were collect. The first 3000 data points were used to estimate the process model, and the last 3000 data points were used for model validation. Three different process models were computed by using the CVA, N4SID and MOESP methods, respectively. One way to validate the estimated model is to insert an excitation sequence, as inputs, to both the process and the estimated model, then compare the process outputs and the model outputs. The block diagram of this model validation method is shown in Figure 2.4, where $\hat{P}$ represents the estimated model, u represents the excitation inputs, y represents the sampled process outputs and $\hat{y}$ represents the model outputs.

The excitation inputs in the last 3000 data points were inserted to the model estimated using the CVA method. The measured process outputs and model outputs are plotted in the same figure. For clarity, only 1000 out of the 3000 data points are shown in Figures 2.5 and 2.6.

The validation results of the model estimated by using the N4SID method are

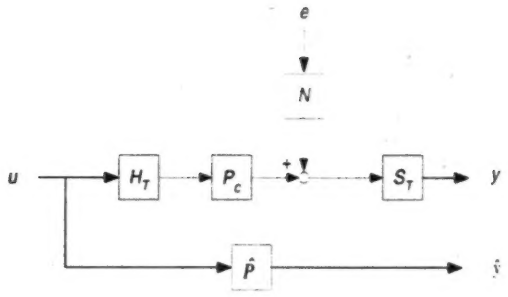


Figure 2.4: *Block diagram for model validation*

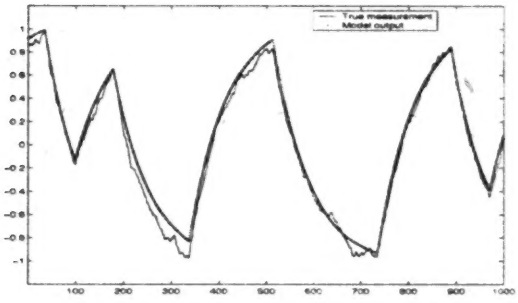


Figure 2.5: *The measured water level and the output of the model estimated using the CVA method*

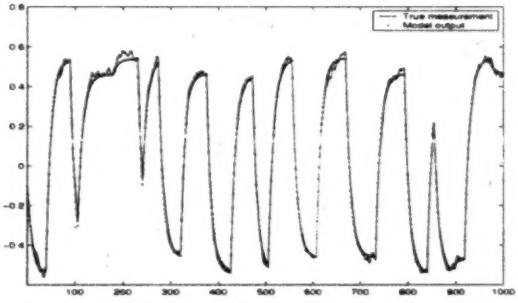


Figure 2.6: *The measured water temperature and the output of the model estimated using the CVA method*

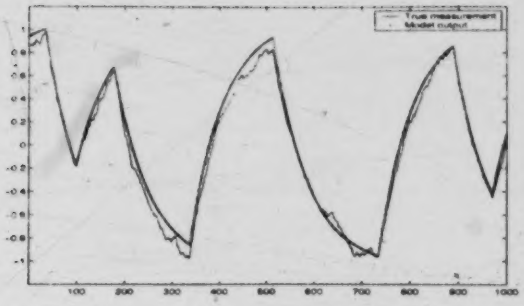


Figure 2.7: *The measured water level and the output of the model estimated using the N4SID method*

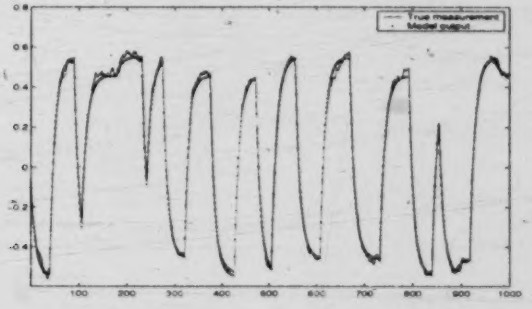


Figure 2.8: *The measured water temperature and the output of the model estimated using the N4SID method*

shown in Figures 2.7 and 2.8.

The outputs of the model estimated using the MOESP approach are compared with the true measurement in Figures 2.9 and 2.10.

2.6.2 Simulation Example

The following model,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{0.1416}{z-0.9865} & 0 \\ \frac{-0.0379}{z-0.913} & \frac{0.0595}{z-0.9048} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$

is used to represent the stirred tank heater in the experiment. A simulation system as shown in Figure 2.2 is generated in SIMULINK. A random binary vector sequence (RBS) was generated as excitation sequences. The signal to noise ratio is 5. A total of 3000 data points were collected. Three different models were estimated by using the CVA, N4SID and MOESP methods, respectively. The step responses of the true

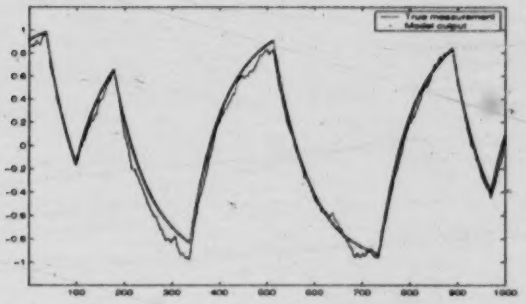


Figure 2.9: *The measured water level and the output of the model estimated using the MOESP method*

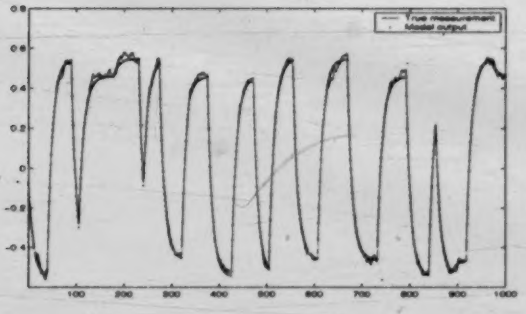


Figure 2.10: *The measured water temperature and the output of the model estimated using the MOESP method*

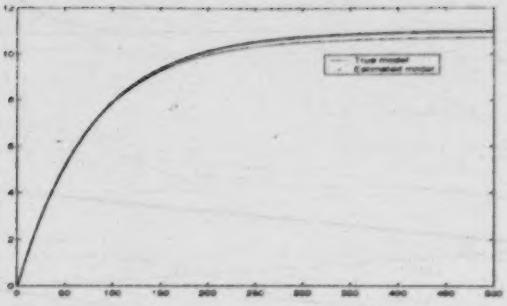


Figure 2.11: *The step responses of the process from u_1 to y_1 and the model estimated using the CVA method*

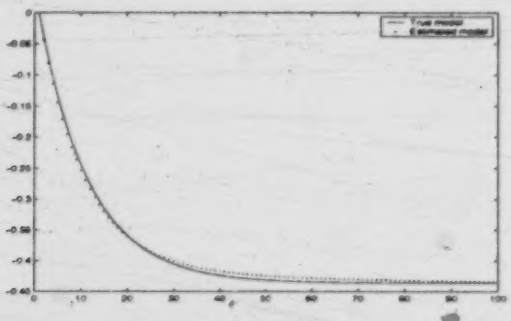


Figure 2.12: *The step responses of the process from u_1 to y_2 and the model estimated by using the CVA method*

models and the estimated models were compared, see Figure2 2.11- 2.19.

The main purpose of this section not to compare these three subspace-based identification methods, but to show whether they work or not. From the experimental and simulation results, we can see that when there is enough excitation, all the subspace-based algorithms can give good results.

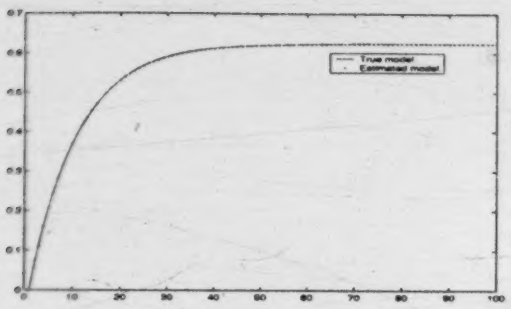


Figure 2.13: *The step responses of the process from u_2 to y_2 and the model estimated by using the CVA method*

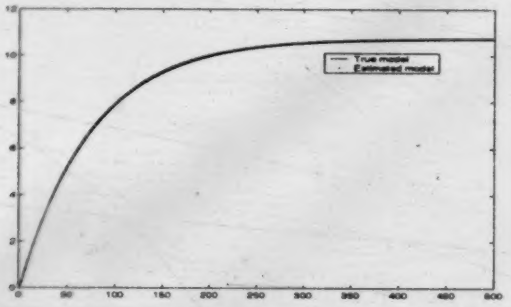


Figure 2.14: *The step responses of the process from u_1 to y_1 and the model estimated by using the N4SID method*

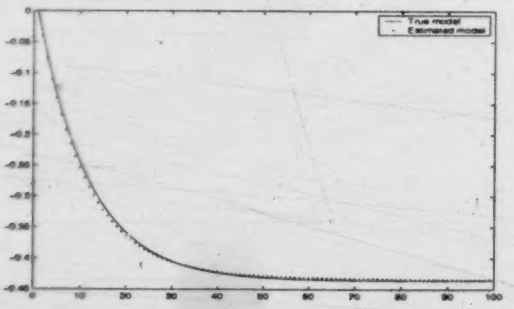


Figure 2.15: *The step responses of the process from u_1 to y_2 and the model estimated by using the N4SID method*

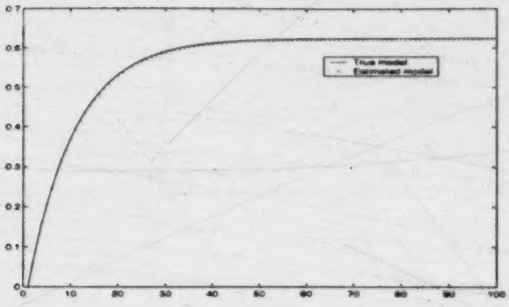


Figure 2.16: *The step responses of the process from u_2 to y_2 and the model estimated by using the N4SID method*

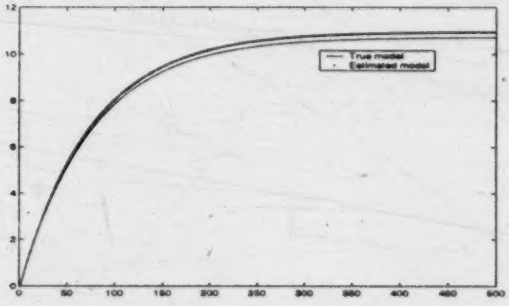


Figure 2.17: *The step responses of the process from u_1 to y_1 and the model estimated by using the MOESP method*

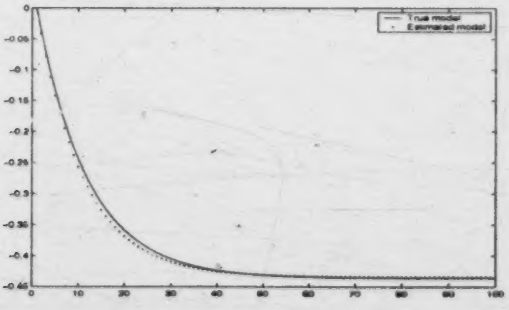


Figure 2.18: *The step responses of the process from u_1 to y_2 and the model estimated by using the MOESP method*

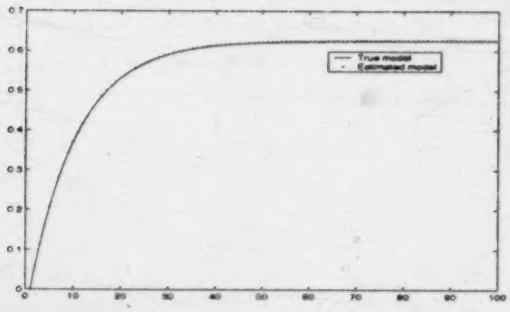


Figure 2.19: *The step responses of the process from u_2 to y_2 and the model estimated by using the MOESP method*

Chapter 3

Introduction to Multirate Systems

Different sampling/hold periods in one system give rise to a multirate system. A simple single-input single-output (SISO) multirate sampled-data system shown in Figure 3.1 is the objective of study in this chapter. In Figure 3.1, u represents the control signal, P_c represents the continuous LTI process, y represents the sampled output, H_T represents a zero-order-hold with period T, and S_{nT} represents an output sampler with period nT. High frequency dots represent fast-rate signals, low frequency dots represent slow-rate signals, and $n(>1)$ is an integer. The multirate system in Figure 3.1 is a multirate system with fast control movement and slow sampling which is typical (e.g., distillation columns, bio-reactors) in chemical industry.

In this chapter, at first the lifting and inverse lifting operators will be introduced and applied to analyzing two classes of multirate systems; then a brief discussion of the properties of the lifting and inverse lifting operators will be given; and finally three different control schemes for the multirate system in Figure 3.1 will be discussed.

3.1 The Lifting Technique and Lifted Systems

Note that in Figure 3.1 both $u(k)$ and $y(k)$ are discrete-time signals defined on the time set $\mathcal{Z}_+ := \{0, 1, 2, \cdots\}$; but their underlying periods are T and nT, respectively. Throughout the chapter we assume P_c is LTI. However, the discrete-time system from $u(k)$ to $y(k)$ is linear periodically time-varying (LPTV). The LPTV nature of multirate systems makes the control theory in the LTI framework inaccessible. In order to get a tractable model for this multirate system, we use the lifting technique. The

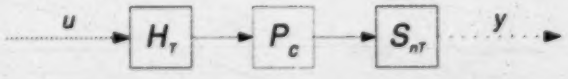


Figure 3.1: *Block diagram of a SISO multirate sampled-data system with fast hold*

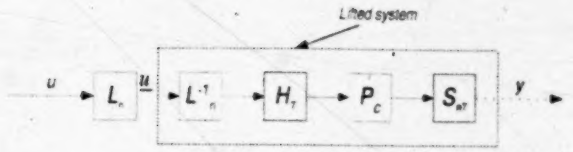


Figure 3.2: *Block diagram of a SISO lifted multirate sampled-data system*

lifting technique was first proposed by Kranc in 1957 [43] as the switch decomposition technique; later, Friedland further developed the concept of lifting which converts a periodic discrete-time system into a time-invariant system [24]; the lifting framework widely accepted now was developed by Khargonekar et al. in 1985 [42].

Let $u(k)$ be a discrete-time signal defined on $\mathcal{Z}_+$. The n-fold lifting operator L_n maps u to $\underline{u}$ (underline denotes lifting) as follows:

$$\{u(0), u(1), u(2), ...\} \mapsto \left\{ \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(n-1) \end{bmatrix}, \begin{bmatrix} u(n) \\ u(n+1) \\ \vdots \\ u(2n-1) \end{bmatrix}, \cdots \right\}.$$

We write $\underline{u} = L_n u$. Note that the dimension of the lifted signal $\underline{u}$ is n times that of u, and the underlying period of $\underline{u}$ again is n times that of u. The inverse process, L_n^{-1}, mapping $\underline{u}$ back to u, is defined as,

$$\left\{ \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(n-1) \end{bmatrix}, \begin{bmatrix} u(n) \\ u(n+1) \\ \vdots \\ u(2n-1) \end{bmatrix}, \cdots \right\} \mapsto \{u(0), u(1), u(2), ...\}.$$

It is easy to see that the following identities hold:

$$L_n^{-1} L_n = I, \quad L_n L_n^{-1} = I.$$

The lifting operator is also norm-preserving:

$$\|Lu\|_2 = \|u\|_2$$

Because of $L_n^{-1} L_n = I$, the multirate system in Figure 3.2 is equivalent to that in Figure 3.1.

$\underline{u} = L_n u$ in Figure 3.2 is the lifted control signal. The fictitious system relating $\underline{u}$ and y is the so called lifted system. Since both $\underline{u}$ and y have the same period, nT, the lifted system is a single-rate system. The lifted system can also be shown to be linear

time-invariant [42]. If the state-space model of the fast-sampled process, $S_T P_c H_T$, is known, then the state-space model of the lifted system can be derived easily: Assume that the state-space model for the system $S_T P_c H_T$ is (A, B, C, D), then the following equations can be derived:

$$\begin{aligned} x(nk+1) &= Ax(nk) + Bu(nk) \\ x(nk+2) &= A^2x(nk) + ABu(nk) + Bu(nk+1) \\ &\vdots \\ x(nk+n) &= A^nx(nk) + A^{n-1}Bu(nk) + ... + Bu(nk+n-1) \\ y_f(nk) &= Cx(nk) + Du(nk), \end{aligned}$$

where y_f represents the fictitious fast-sampled output. Define

$$\begin{aligned} \underline{x}(k) &\triangleq x(nk), \ \underline{x}(k+1) \triangleq x(nk+n) \\ u_1(k) &\triangleq u(nk), \ u_2(k) \triangleq u(nk+1), \ ..., \ u_n(k) \triangleq u(nk+n-1) \\ y(k) &\triangleq y_f(nk). \end{aligned}$$

Let $\underline{P}$ represent the lifted system, $\underline{x}$ the state of the lifted system, $\underline{u} = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix}^T$ the input to this lifted system, and y the output of the lifted system. The state-space model of the lifted system can then be written as

$$\begin{aligned} \underline{x}(k+1) &= A^n\underline{x}(k) + A^{n-1}Bu_1(k) + ... + Bu_n(k) \\ y(k) &= C\underline{x}(k) + \begin{bmatrix} D & 0 & \cdots & 0 \end{bmatrix}\underline{u}. \end{aligned}$$

The state-space matrices $(\underline{A}, \underline{B}, \underline{C}, \underline{D})$ for the lifted system, $\underline{P}$, are given by

$$\left[\begin{array}{c|c} \underline{A} & \underline{B} \\ \hline \underline{C} & \underline{D} \end{array}\right] = \left[\begin{array}{c|cccc} A^n & A^{n-1}B & A^{n-2}B & \cdots & B \\ \hline C & D & 0 & \cdots & 0 \end{array}\right]. \tag{3.1}$$

Though the continuous process is a SISO process, the lifted system is a multi-input and single-output (MISO) system. Clearly, the lifting operation results in an increased input-output dimensionality. Multirate systems with fast control rates and slow output sampling rates are the common pattern in chemical industry, and this class of multirate systems is the focus of the study in this chapter. However, for the sake of completeness, the case where control is slow-rate and output sampling is fast-rate will be discussed in the rest of this section although it is rare in chemical industry.

For simplicity, we assume that the control period is an integer (n) multiple of the output sampling period. For some reasons, the control interval can only be relatively

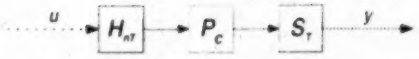


Figure 3.3: *Block diagram of a SISO multirate sampled-data system with slow hold*

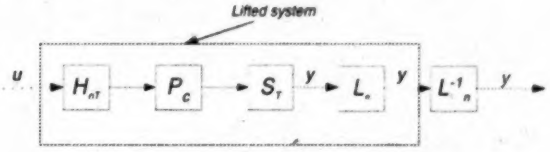


Figure 3.4: *Block diagram of a lifted SISO multirate sampled-data system with slow hold*

large, but at the same time the output can be sampled at relatively fast rate. For such multirate systems it will be helpful if the inter-sample outputs are considered in the controller design. In order to obtain an LTI system which relates the slow-sampled input to all the fast-sampled outputs, the lifting technique is applied to the multirate signals shown in Figure 3.3. This results in the system shown in Figure 3.4.

The state-space model of the lifted system can be derived from the state-space model of the fast-sampled system: The following equations can be derived from the state-space model of the fast-sampled system:

$$\begin{aligned}
x(nk+1) &= Ax(nk) + Bu_f(nk) \\
x(nk+2) &= A^2x(nk) + ABu_f(nk) + Bu_f(nk+1) \\
&\vdots \\
x(nk+n) &= A^nx(nk) + A^{n-1}Bu_f(nk) + ... + Bu_f(nk+n-1) \\
y(nk) &= Cx(nk) + Du_f(nk) \\
y(nk+1) &= CAx(nk) + CBu_f(nk) + Du_f(nk+1) \\
&\vdots \\
y(nk+n-1) &= CA^{n-1}x(nk) + CA^{n-2}Bu_f(nk) + CA^{n-3}Bu_f(nk+1) + \cdots + Du_f(nk+n-
\end{aligned}$$

where u_f represents the fictitious fast-sampled control movement. During interval nT, the control signal is constant, so

$$u_f(nk) = u_f(nk+1) = \cdots = u_f(nk+n-1) = u(k).$$

Define

$$\begin{aligned}
\underline{x}(k) &\triangleq x(nk), \quad \underline{x}(k+1) \triangleq x(nk+n), \\
y_1(k) &\triangleq y(nk),\ y_2(k) \triangleq y(nk+1), \cdots,\ y_n(k) \triangleq y(nk+n-1).
\end{aligned}$$

Let $\underline{x}$ represent the state of the lifted system, u the input to this lifted system, and $\underline{y} = \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^T$ the output of the lifted system. The state-space model of the lifted system can then be written as:

$$\begin{aligned} \underline{x}(k+1) &= A^n \underline{x}(k) + (A^{n-1}B + ... + B)u(k) \\ y(k) &= \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \underline{x}(k) + \begin{bmatrix} D \\ CB + D \\ \cdots \\ CA^{n-2}B + CA^{n-3}B + \cdots + D \end{bmatrix} u(k). \end{aligned}$$

The state-space model of the lifted system, $\underline{P}$, is given by

$$\left[\begin{array}{c|c} A^n & A^{n-1}B + ... + B \\ \hline C & D \\ CA & CB + D \\ \vdots & \cdots \\ CA^{n-1} & CA^{n-2}B + CA^{n-3}B + \cdots + D \end{array}\right].$$

3.2 Multirate Control

Since discrete-time multirate systems from $u(k)$ to $y(k)$ are linear periodically time-varying, the controller design for multirate systems is more complicated than that of the single-rate systems. In this section, we will use the multirate system in Figure 3.1 as an example to illustrate three control strategies for multirate systems.

3.2.1 Slow-Rate Control

A possible way to control the multirate system in Figure 3.1 is to design a single-rate controller, K_s, with control interval nT, as shown in Figure 3.5. For a general MIMO multirate system, we can design a single-rate controller with an interval which is the least common multiple of all the output sampling intervals and hold periods. The advantage of this strategy is that it is simple because now the system becomes a single-rate system. But in this case, the capacity of the equipment is not fully explored; and the inter-sample performance may be very poor.

3.2.2 Lifted Control

Lifting techniques are powerful tools in analyzing multirate systems because of their capability of converting multirate systems into single-rate LTI systems. Another strategy to overcome the LPTV drawback is to design a lifted control scheme. The

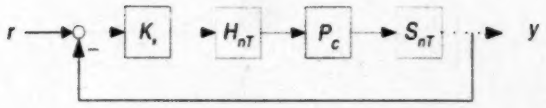


Figure 3.5: *Block diagram of a SISO multirate sampled-data system with a slow SR controller*

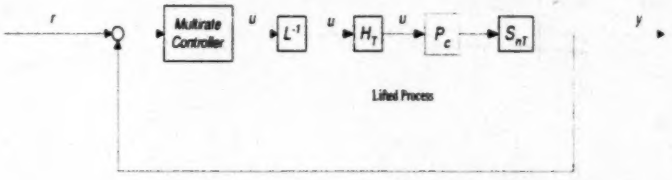


Figure 3.6: *Block diagram of a SISO multirate sampled-data system with a multirate controller*

lifted control scheme has three steps. The first step is to lift all the input-output signals with different sampling rates to fictitious signals at one uniform sampling rate. This step results in fictitious lifted systems with increased input-output dimensionality. The second step involves designing a controller for the lifted system. Since the output of the lifted controller is the lifted control moves, the lifted control moves have to be inverse lifted to obtain the fast-rate input signals to the process. The fundamental idea of the lifted control scheme can be illustrated by designing a lifted control scheme for the simple multirate system in Figure 3.1; this results in a multirate control system as shown in Figure 3.6. Numerous advanced control algorithms using the lifted control strategy have been developed for several classes of multirate systems [9, 11, 66] in the last decade.

The idea of the lifted control strategy is simple, and the lifted control scheme can achieve better performance compared with the slow-rate control strategy, but there are a few constraints in the lifted control strategy. The first constraint is the causality constraint. We will use a simple example to illustrate the causality constraint problem: Consider a SISO multirate system where the control interval is $2T$ and the output sampling interval is $3T$. In order to obtain the lifted system, the input and output signals have to be lifted as follows:

$$\underline{u}(0) = \begin{bmatrix} u(0) \\ u(2T) \\ u(4T) \end{bmatrix}, \quad \underline{y}(0) = \begin{bmatrix} y(0) \\ y(3T) \end{bmatrix}.$$

Assume that the reference signal is zero, then the controller designed for the lifted

system should be

$$\underline{u} = -\begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \\ G_{31} & G_{32} \end{bmatrix} \underline{y}.$$

It is evident that at time instant 0, output $y(3T)$ is unknown; and at time instant $2T$, output $y(3T)$ is also unknown. So G_{21} and G_{22} should have at least one sample time-delay otherwise the past control moves, $u(0)$ and $u(2T)$, would depend on the future output, $y(3T)$, and clearly such controller is not causal.

Another constraint is the gain constraint. In order to avoid the inter-sample ripples in the fast-sampled outputs, the gains of the lifted controller should satisfy certain constraint [71]. We can use the multirate system in Figure 3.1 to briefly illustrate the gain constraint problem. Let $\underline{G}$ represent the multirate controller in Figure 3.6. Clearly $\underline{G}$ should have such structure:

$$\underline{G} = \begin{bmatrix} G_1 \\ \vdots \\ G_n \end{bmatrix}.$$

For simplicity, we assume here that G_i, $i = 1, ..., n$, are all constants (This means that all the controllers in the lifted controller are proportional controllers); and the closed-loop system with lifted controller is also stable. When the closed-loop system reaches steady-state, at sampling interval T the inputs to the process are $u(nkT+(i-1)T) = G_i e(\infty), i = 1, 2, ..., n$, where $e(\infty)$ is the steady-state control error. In order to avoid the inter-sample ripples, $u(nkT) = u(nkT + T) = ... = u(nkT + (n-1)T)$ should be satisfied. This means that the gains of G_i, $i = 1, ..., n$ (since $G_i, i = 1, ..., n$ are constants, $G_i, i = 1, ..., n$ are also the gains of G_i), should be uniform in order to avoid inter-sample ripples. More detailed analysis can be found in [71].

It is evident that these constraints make the lifted control schemes complicated to design. Moreover it is also more difficult to analyze the lifted control systems due to the increased dimensionality.

3.2.3 Inferential Control

As explained in Section 1.1, if the unmeasured inter-sample outputs are estimated, then the estimated inter-sample outputs can be used for the control purpose. This is the principle of inferential control schemes, and clearly inter-sample estimation is the key to inferential control schemes. All the inter-sample estimation methods can be classified into two cases: model-based inter-sample estimation and non-model-based inter-sample estimation. The inference model can be, for example, the fast-sampled model of the process, the empirical model relating the primary variables and

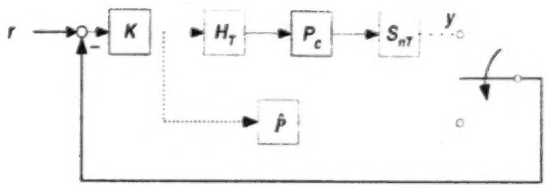


Figure 3.7: *Block diagram of a SISO multirate sampled-data system with an inferential controller*

the secondary variables, etc.. The empirical model-based inferential control schemes have been practiced in chemical industry (e.g., distillation columns, polymerization reactors) for a long time. The fast-sampled model of a process can be obtained by either fast discretization of a continuous-time model or estimating from multirate data. In most cases, it is impossible to know the continuous-time models. However it is feasible to estimate the fast-sampled model from multirate data only after a two-step method developed by Li et. al in 1999 [50].

Figure 3.7 shows the block diagram of the multirate system in Figure 3.1 with the inferential control scheme. In Figure 3.7, $\hat{P}$ represents the estimated fast-sampled model, the switch connects with the slow output sampler every period, nT, and connects with the fast-sampled model during the inter-sample instants $(nT+T, nT+2T, ..., nT+(n-1)T)$. Fast-sampled model-based inferential control schemes, as shown in Figure 3.7, are applied to chemical industry only recently [51]. Details of this industrial application will be discussed in Chapter 8. The advantage of the inferential control scheme is that it is simple since all the single-rate control techniques can be applied to it. Compared with the slow-rate control scheme, the inferential control scheme can achieve better performance since it provides fast-rate control moves. The inferential control scheme does not have such constraints as the lifted control scheme has, and the inferential control scheme also enjoys smaller dimensionality compared with the lifted control scheme. The disadvantage of the inferential control scheme is that it requires a fast-sampled process model, but fortunately, this disadvantage has been overcome by the two-step estimation method proposed by Li et. al [49].

Chapter 4

Subspace-based Identification of SISO Multirate systems

4.1 Introduction

In Chapter 3, we introduced three methods for controlling the multirate systems. The last two methods, lifted control strategy and the inferential control strategy, are more popular, since they provide better performance compared with the slow-rate control strategy. The lifted model-based predictive control strategy is based on the lifted model; and the inferential control strategy always requires the fast-sampled process model. It is quite often that the the input-output measurement is the only source to obtain the model. The objective of this chapter is to develop a practical method to estimate the lifted model and the fast-sampled model from the input-output data.

We know that even if a processes only have single input and single output, the lifted system is a MISO or MIMO system. Evidently it is very efficient to represent a MIMO system with a state-space model. So it is a natural choice to identify the lifted system in the state-space framework. Subspace based system identification is relatively new, but it has achieved tremendous success since it was first developed at the beginning of the last decade. A number of subspace identification algorithms, for example the CVA, N4SID and MOESP etc. have been published. Compared with the identification algorithms in the polynomial framework, the subspace based identification methods are non-iterative, with no non-linear optimization part involved [77], and numerically more stable.

Identifying the dynamic models of multirate systems is still an relatively new area, not much research work has been done in this field. Continuous-time identification methods [73, 25, 33] have drawn much attention recently; but the identified continuous model is always biased in the presence of noise [25, 33] except when the output is

corrupted only by white noise [14]. It is well known that based on sampled data, one cannot uniquely identify the continuous-time model. Verhaegen and Yu presented a technique to estimate lifted models of multirate systems in the state-space domain [82]. In their work, they represented multirate systems with periodic systems which are difficult to be understood by engineers, and they could estimate the lifted model with only the MOESP approach. Our objective is to develop a strategy which can apply most of the existing subspace-based identification algorithms to estimating the lifted models. In this chapter we will focus on a general class of single-input and single-output multirate systems where the ratio between the output sampling interval and control interval is a rational number; and present a two-step multirate identification algorithm: estimate the lifted process model and then extract a fast sampled model from the lifted model.

This chapter is organized as follows. In Section 2 we present the lifting technique and lifted multirate systems in state-space forms. In Section 3 we show that in the presence of time delays, observability of lifted systems may be lost, depending on how large the time delays are. In Section 4 we give a method for estimating the time delay of the process from the lifted model, and analyze the accuracy of this estimation process. In Section 5 we discuss the causality issues involved in the lifted models and propose a subspace-based algorithm for identifying the state-space models of the lifted systems, respecting the causality constraint. In Section 6 we present two methods for extracting fast-rate models based on the lifted ones. Section 7 summarizes a procedure we propose for identifying a fast-rate model from the multirate data, based on the results studied in the preceding sections. We illustrate and validate the results of this chapter in Section 8 on a simulation example. Finally, the concluding marks are given in Section 9.

4.2 Lifted Systems

Consider the multirate sampled-data system in Figure 4.1. Here, P_c is a continuous-time process with additive noise; the noise is generated by a continuous-time model N with a white noise input e; the output of P_c is corrupted by that of N, and is sampled by a sampler S_{nT} with period nT, yielding the sampled output $y(k)$; the input to P_c is generated by a zero-order hold with period mT processing the input sequence $u(k)$. Without loss of generality, we assume that the two integers m and n are co-prime, for otherwise, we can absorb any common factor of m and n into T, a positive real number called the base period. As mentioned before, such multirate systems arise often in industry; the practically interesting case is when $m < n$, where

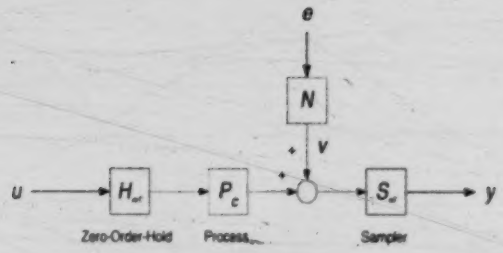


Figure 4.1: *Block diagram of a SISO multirate sampled-data system*

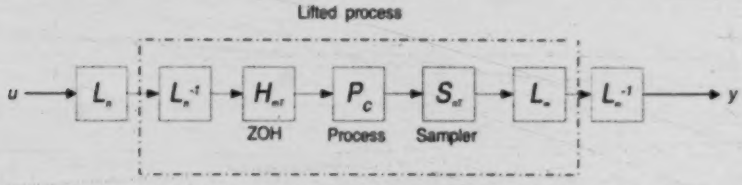


Figure 4.2: *The lifted SISO multirate system*

the control rate is faster than the sampling rate. From now on, we will focus on this case.

Note that both $u(k)$ and $y(k)$ are discrete-time signals defined on the time set $\mathcal{Z}_+ := \{0, 1, 2, \cdots\}$; but their underlying periods are mT and nT, respectively. Throughout the chapter we assume P_c is LTI. However, the discrete-time system from $u(k)$ to $y(k)$, with $e = 0$, is linear periodically time-varying. In order to get a tractable model for this multirate system, we use the lifting technique.

Let $u(k)$ be a discrete-time signal defined on $\mathcal{Z}_+$. The n-fold lifting operator L_n maps u to $\underline{u}$ (underline denotes lifting) as follows:

$$\{u(0), u(1), u(2), ...\} \mapsto \left\{ \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(n-1) \end{bmatrix}, \begin{bmatrix} u(n) \\ u(n+1) \\ \vdots \\ u(2n-1) \end{bmatrix}, \cdots \right\}.$$

We write $\underline{u} = L_n u$. Note that the dimension of the lifted signal $\underline{u}$ is n times that of u, and the underlying period of $\underline{u}$ again is n times that of u. For the multirate system from u to y in Figure 4.1, we intend to lift the input and output to get a single-rate system with underlying period mnT; and hence we need to lift u by L_n and y by L_m to arrive at Figure 4.2, where the lifted multirate system $\underline{P}$ takes $\underline{u} = L_n u$ into $\underline{y} = L_m y$ and is defined via

$$\underline{P} = L_m S_{nT} P_c H_{mT} L_n^{-1}. \tag{4.1}$$

This $\underline{P}$ can be shown to be LTI. In order to derive a state-space model for $\underline{P}$, we discretize P_c via the zero-order hold method with the sampling period T to get $P_T := S_T P_c H_T$, S_T and H_T being the sampler and zero-order hold with period T, and assume that P_T has a state-space model

$$D + C(zI - A)^{-1}B =: \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]. \tag{4.2}$$

It is not hard to derive that a state-space model for the discretization with period mT, $P_{mT} = S_{mT} P_c H_{mT}$, is given by

$$P_{mT}(z) = \left[\begin{array}{c|c} A_{mT} & B_{mT} \\ \hline C & D \end{array}\right] \tag{4.3}$$

with

$$A_{mT} = A^m, \quad B_{mT} = (I + A + \cdots + A^{m-1})B. \tag{4.4}$$

Proposition 1 *A state-space model for the lifted multirate system $\underline{P}$ is given by*

$$\underline{P}(z) = \left[\begin{array}{c|ccc} A_{mT}^n & A_{mT}^{n-1}B_{mT} & \cdots & B_{mT} \\ \hline C & D & \cdots & 0 \\ CA^n & CA^{n-m}B_{mT} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ CA^{(m-1)n} & CA^{mn-n-m}B_{mT} & \cdots & 0 \end{array}\right]. \tag{4.5}$$

Observe that certain blocks in the D-matrix are zero; this reflects the causality constraint.

Proof of Proposition 1 From (4.1) and the identities $S_{nT} = S_{nT} H_T S_T$ and $H_{mT} = H_T S_T H_{mT}$, we can write

$$\begin{aligned} \underline{P} &= L_m S_{nT} H_T (S_T P_c H_T) S_T H_{mT} L_n^{-1} \\ &= L_m S_{nT} H_T L_{mn}^{-1} L_{mn} P_T L_{mn}^{-1} L_{mn} S_T H_{mT} L_n^{-1} \\ &= \underline{S}\,\underline{P}_T\,\underline{H}, \end{aligned}$$

the last equality following from the definitions:

$$\underline{S} = L_m S_{nT} H_T L_{mn}^{-1}, \quad \underline{P}_T = L_{mn} P_T L_{mn}^{-1}, \quad \underline{H} = L_{mn} S_T H_{mT} L_n^{-1}.$$

Now based on the state-space model of P_T in (4.2), a state-space model for $\underline{P}_T$ is given by [42]:

$$\underline{P}_T(z) = \left[\begin{array}{c|cccc} A^{mn} & A^{mn-1}B & A^{mn-2}B & \cdots & B \\ \hline C & D & 0 & \cdots & 0 \\ CA & CB & D & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ CA^{mn-1} & CA^{mn-2}B & CA^{mn-3}B & \cdots & D \end{array}\right]. \tag{4.6}$$

Also, it can be shown that $\underline{S}$ and $\underline{H}$ are constant matrices given by

$$\underline{S} = \begin{bmatrix} \overbrace{I\ 0\ \dots\ 0}^{n} & \overbrace{0\ 0\ \dots\ 0}^{n} & \dots & \overbrace{0\ 0\ \dots\ 0}^{n} \\ 0\ 0\ \dots\ 0 & I\ 0\ \dots\ 0 & \dots & 0\ 0\ \dots\ 0 \\ \vdots\ \vdots\ \vdots\ \vdots & \vdots\ \vdots\ \vdots\ \vdots & \vdots & \vdots\ \vdots\ \vdots\ \vdots \\ 0\ 0\ \dots\ 0 & 0\ 0\ \dots\ 0 & \dots & I\ 0\ \dots\ 0 \end{bmatrix}_{m\times(mn)\text{ blocks}}$$

$$\underline{H} = \begin{bmatrix} I & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ I & 0 & 0 & 0 & \cdots & 0 \\ 0 & I & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & I & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & I \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & I \end{bmatrix} \begin{matrix} \\ \\ m \\ \\ \\ m \\ \\ \\ m \\ \\ \\ \end{matrix}_{(mn)\times n\text{ blocks}}$$

(Here, the identity matrices I reduce to 1 if P is a SISO process.) Pre- and post-multiplying the transfer matrix in (4.6) by $\underline{S}$ and $\underline{H}$, respectively, gives a state-space model for $\underline{P}$:

$$\left[\begin{array}{c|ccc} A^{mn} & (\sum_{i=mn-m}^{mn-1} A^i)B & (\sum_{i=mn-2m}^{mn-m-1} A^i)B & \cdots & (\sum_{i=0}^{m-1} A^i)B \\ \hline C & D & 0 & \cdots & 0 \\ CA^n & C(\sum_{i=n-m}^{n-1} A^i)B & D + C(\sum_{i=0}^{n-m-1} A^i)B & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ CA^{(m-1)n} & C(\sum_{i=mn-n-m}^{mn-n-1} A^i)B & C(\sum_{i=mn-n-2m}^{mn-n-m-1} A^i)B & \cdots & 0 \end{array}\right].$$

This model simplifies to (4.5) by noting (4.4). ∎

As a special case, when $m = 1$, the state-space model for $\underline{P}$ in Proposition 1 simplifies to

$$\left[\begin{array}{c|cccc} A^n & A^{n-1}B & A^{n-2}B & \cdots & B \\ \hline C & D & 0 & \cdots & 0 \end{array}\right]. \tag{4.7}$$

Lifting the noise model is simpler and so we only briefly discuss the result. Discretize the continuous-time noise model N in Figure 4.1 with sampling period nT to get N_{nT} which takes a white noise $e(k)$ into $v(k)$; assume a state-space model

$$N_{nT}(z) = \left[\begin{array}{c|c} A_N & B_N \\ \hline C_N & D_N \end{array}\right].$$

Since the deterministic system is lifted to interval mnT, we lift N_{nT} by L_m: Define $\underline{v} = L_m v$ and $\underline{e} = L_m e$ to get that the lifted noise model, $\underline{N} := L_m N_{nT} L_m^{-1}$, maps $\underline{e}$ to $\underline{v}$ and admits a state-space model

$$\underline{N}(z) = \left[\begin{array}{c|ccc} A_N^m & A_N^{m-1}B_N & A_N^{m-2}B_N & \cdots & B_N \\ \hline C_N & D_N & 0 & \cdots & 0 \\ C_N A_N & C_N B_N & D_N & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ C_N A_N^{m-1} & C_N A_N^{m-2} B_N & C_N A_N^{m-3} B_N & \cdots & D_N \end{array}\right].$$

Thus we get the overall lifted model as follows:

$$\underline{y} = \underline{P}\,\underline{u} + \underline{N}\,\underline{e}. \tag{4.8}$$

Both $\underline{P}$ and $\underline{N}$ are now LTI, but $\underline{u}$ and $\underline{e}$ have higher dimensions. Note that most statistical properties of $e(k)$ are preserved after lifting: If $e(k)$ is a white noise, so is $\underline{e}(k)$; if $e(k)$ is Gaussian, so is $\underline{e}(k)$; and so on.

4.3 Observability in the Presence of Time Delay

When is the state-space model for the lifted multirate system in (4.5) controllable and observable? To answer this question, our standing assumption in this section is that the state-space model for P_T in (4.2) is minimal, i.e., (A, B) is controllable and (C, A) is observable. Note that this assumption is guaranteed if the continuous-time process with no time delay is controllable and observable and if the sampling period T is non-pathological [10]. Controllability of $\underline{P}$ in (4.5) can be achieved under a mild condition [10, 23]; so let us focus on observability. We need the following sufficient condition.

Condition 1 *For every eigenvalue λ of A, none of the $mn-1$ points*

$$\lambda e^{2\pi kj/mn}, \quad k = 1, 2, \cdots, mn-1,$$

is an eigenvalue of A.

Under Condition 1, observability of (C, A) implies that of (C, A^{mn}) [23], which in turn implies observability of $\underline{P}$ in (4.5).

Now we restrict our attention to P_c being SISO. In the time-delay case, let the time delay in P_c be $d_c T$, d_c being a real positive number. Suppose $d_c > 1$. It follows that the discretized system P_T has at least two poles at $z = 0$; and hence A has at least two eigenvalues at $z = 0$. Thus Condition 1 fails in this case. We will prove that surprisingly, we lose observability of the model in (4.5).

Proposition 2 *The state-space model in (4.5) for $\underline{P}$ is unobservable if $d_c > 1$; it is observable if $0 \le d_c \le 1$ and if Condition 1 is satisfied.*

Proof The second part follows readily from [23. 10]. We now prove the first part.

Let d be the smallest integer such that $d_c \le d$. It follows that $d \ge 2$ since $d_c > 1$. The discretized system P_T has d poles at $z = 0$. Without loss of generality, we can assume that the state-space model of P_T in (4.2) is in the controllable canonical form; in this case, the A-matrix is of the form

$$A = \overbrace{\begin{bmatrix} a_1 & a_2 & \cdots & a_l & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}}^{l \qquad\qquad d} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix},$$

where the dimensions of A_{11} and A_{22} are $l \times l$ and $d \times d$, respectively. It follows easily that for $k \ge 2$, the last two columns of A_{22}^k are zero; so are those of A^k.

Now we express the observability matrix for the state-space model in (4.5) in terms of A and C:

$$\Gamma_o = \begin{bmatrix} C \\ CA^n \\ CA^{2n} \\ \vdots \\ CA^{mn(d+l)-n} \end{bmatrix}$$

Because $n > m \ge 1$, every row in Γ_o except the first one has at least two zero elements at the end; so the rank of Γ_o is at most $l + d - 1$. Therefore we loose observability. ■

Loss of observability would cause problems for some applications of the identified lifted models, e.g., making it impossible to extract fast single-rate models from the identified lifted models. In order to understand the time-delay related observability problem clearly, let us see the following example: There is a single-input single-output system with control interval 1 unit of time and output sampling interval 3 units of time (that is the special case where $m = 1$). The single-rate process model with sampling interval 1 unit of time is $\frac{0.1}{z^3 - 0.9}$. The state-space model with sampling interval 1 unit of time is

$$A = \begin{bmatrix} 0.9 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \; B = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$
$$C = \begin{bmatrix} 0 & 0 & 0.1 \end{bmatrix}, \; D = 0.$$

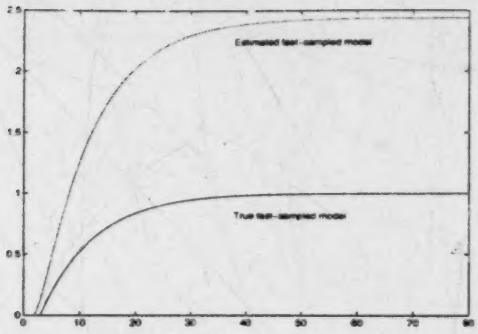


Figure 4.3: *Comparison between the true fast model and the fast model extracted by using the controllability approach*

The lifted model in the form of equation (4.7) is

$$\begin{aligned} \underline{A} &= \begin{bmatrix} 0.729 & 0 & 0 \\ 0.81 & 0 & 0 \\ 0.9 & 0 & 0 \end{bmatrix}, \underline{B} = \begin{bmatrix} 0.81 & 0.9 & 1 \\ 0.9 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \\ \underline{C} &= \begin{bmatrix} 0 & 0 & 0.1 \end{bmatrix}, \underline{D} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}. \end{aligned} \tag{4.9}$$

We can see that the pair $(\underline{C}, \underline{A})$ is not observable. The minimum realization form of the lifted model is

$$\begin{aligned} \underline{A} &= \begin{bmatrix} 0.729 & 0 \\ 0.9 & 0 \end{bmatrix}, \underline{B} = \begin{bmatrix} 0.81 & 0.9 & 1 \\ 1 & 0 & 0 \end{bmatrix} \\ \underline{C} &= \begin{bmatrix} 0 & 0.1 \end{bmatrix}, \underline{D} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}. \end{aligned} \tag{4.10}$$

Since the lifted model in (4.9) is not observable, we could only estimate the lifted model in (4.10). From the lifted model in (4.10), we can extract fast-sampled models (model with sampling interval 1 unit of time) by using the methods in Section 4.6 (we could not use the eigenvalue approach since an eigenvalue of $\underline{A}$ is zero). The extracted fast-sampled models are then compared with the true fast-sampled model in Figures 4.3 and 4.4. Clearly they do not agree.

From the comparison we can see that we could not obtain the true fast-sampled model if observability of the lifted model in 4.7 is lost. One remedy is as follows: In the next section we will show that based on the multirate input-output data, we can estimate the time delay $d_c T$ with accuracy T; say, the estimated time delay is $d_d T$ with d_d an integer and $0 < d_c - d_d \leq 1$. Since m and n are co-prime, we can always find integers k_1 and k_2 such that

$$d_d = k_1 m + k_2 n.$$

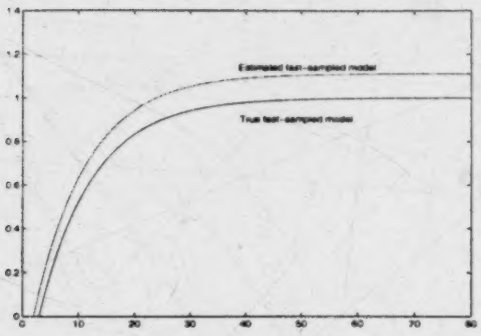


Figure 4.4: *Comparison between the true fast model and the fast model extracted by using the matrix roots approach*

Next we shift the input data to the left by k_1 samples and the output data to the right by k_2 samples, so that the shifted input-output data gives rise to a new system with time delay no larger than T. The lifted model identified from the shifted data will be observable under a mild condition (Proposition 2).

4.4 Time-Delay Estimation

In this section we study how to estimate the time delay (d_cT) of the continuous-time process P_c using multirate data. If (discrete) time delays for the subsystems of the lifted multirate process $\underline{P}$ can be estimated – a standard problem, we will show a method for estimating d_cT with accuracy T.

The lifted process model with noise is given in (4.8). Suppose P_c is SISO; the lifted process $\underline{P}$ has a $m \times n$ transfer matrix:

$$\underline{P}(z) = \begin{bmatrix} P_{00}(z) & P_{01}(z) & \cdots & P_{0,n-1}(z) \\ P_{10}(z) & P_{11}(z) & \cdots & P_{1,n-1}(z) \\ \vdots & \vdots & & \vdots \\ P_{m-1,0}(z) & P_{m-1,1}(z) & \cdots & P_{m-1,n-1}(z) \end{bmatrix} \tag{4.11}$$

Time delays in $P_{ij}(z)$ can be estimated by identifying the impulse response matrix for $\underline{P}$ using the standard correlation analysis between $\underline{u}$ and $\underline{y}$ [55, 68], which we summarize below:

1. Design a vector input sequence $\underline{u}(k)$ whose components are independent to each other and satisfy proper excitation requirement [55]; inverse lift $\underline{u}(k)$ by L_n^{-1} and input to the multirate process; measure the output $y(k)$ and lift by L_m to get $\underline{y}(k)$. (Measurement noise is assumed to be white and independent of $\underline{u}$.)

2. Compute

$$R_{yu}(i) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \underline{y}(k+i)\underline{u}^T(k),$$

$$R_{uu}(i) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \underline{u}(k+i)\underline{u}^T(k),$$

where superscript T means transpose. These are usually approximated by sufficiently large N.

3. If $\underline{u}(k)$ is a vector white noise, $R_{uu}(0)$ is an invertible matrix and $R_{uu}(i) = 0$ for $i \neq 0$. The impulse response matrix for $\underline{P}$ is then calculated to be

$$\underline{p}(i) = R_{yu}(i)R_{uu}^{-1}(0).$$

Now we assume that time delays for the subsystems of $\underline{P}$ in (4.11) are computed: $P_{ij}(z)$ contains a time delay of l_{ij}, l_{ij} being a nonnegative integer. (Thus the first non-zero element in the impulse response of P_{ij} is $p(l_{ij})$.) Knowing such a time-delay matrix:

$$\begin{bmatrix} l_{00} & l_{01} & \cdots & l_{0,n-1} \\ l_{10} & l_{11} & \cdots & l_{1,n-1} \\ \vdots & \vdots & & \vdots \\ l_{m-1,0} & l_{m-1,1} & \cdots & l_{m-1,n-1} \end{bmatrix}, \tag{4.12}$$

how do we estimate d_cT?

The lifted system $\underline{P}$ maps $\underline{u}$ to $\underline{y}$, both having period mnT; and the subsystem P_{ij} maps $\underline{u}_j$ to $\underline{y}_i$. Note that during the first interval $[0, mnT)$, $\underline{u}_j$ occurs at time $j(mT)$ and $\underline{y}_i$ at time $i(nT)$; so the actual time delay from $\underline{u}_j$ to $\underline{y}_i$, incorporating that due to lifting, is

$$T_{ij} = d_cT + jmT - inT. \tag{4.13}$$

From l_{ij}, we can only estimate T_{ij} within one sample period which is mnT for the lifted system:

$$(l_{ij} - 1)mnT < T_{ij} \leq l_{ij}mnT.$$

Substituting (4.13) into the above, we get mn inequalities for estimating d_c:

$$(l_{ij} - 1)mn + in - jm < d_c \leq l_{ij}mn + in - jm, \qquad i = 0, 1, \cdots, m-1, \quad j = 0, 1, \cdots, n-1. \tag{4.14}$$

These inequalities combined give much better accuracy for estimating d_c. To this end, we need the following claim based on the co-primeness of m and n.

Claim *As i goes from 0 to $m-1$ and j from 0 to $n-1$, $in-jm$ takes mn different integer values.*

Proof Suppose, for a contradiction, that there exist (i_1, j_1) and (i_2, j_2) with $0 \leq i_1, i_2 \leq m-1$ and $0 \leq j_1, j_2 \leq n-1$ such that

$$i_1 n - j_1 m = i_2 n - j_2 m.$$

Then

$$(i_1 - i_2)n = (j_1 - j_2)m.$$

The co-primeness of m and n and the inequalities $|i_1 - i_2| < m$ and $|j_1 - j_2| < n$ imply immediately that $i_1 - i_2 = 0$ and $j_1 - j_2 = 0$. Hence the claim is proven. ∎

Over an interval of $2mn$ integers, based on the claim, the mn lower limits in (4.14) are all different and fill the lower half of the interval, whereas the mn upper limits in (4.14) fill the upper half. Thus the difference between the maximum of the lower limits and the minimum of the upper limits is only 1 – we can estimate d_c with accuracy 1 and hence $d_c T$ with accuracy T!

Let us illustrate this with a concrete example. Take $m = 2$ and $n = 3$; assume the time delay of the continuous-time process is $8.5T$, i.e., $d_c = 8.5$. Based on these, the time delay matrix as in (4.12) should be

$$\begin{bmatrix} l_{00} & l_{01} & l_{02} \\ l_{10} & l_{11} & l_{12} \end{bmatrix} = \begin{bmatrix} 2 & 2 & 3 \\ 1 & 2 & 2 \end{bmatrix}.$$

Inequalities in (4.14) reduce to the following:

$$\begin{array}{llll} 6 < d_c \leq 12, & 4 < d_c \leq 10, & 8 < d_c \leq 14, \\ 3 < d_c \leq 9, & 7 < d_c \leq 13, & 5 < d_c \leq 11. \end{array}$$

These six inequalities are equivalent to $8 < d_c \leq 9$; hence we can estimate the time delay within an interval of T: $8T < d_c T \leq 9T$.

4.5 Identification of Lifted Models

The lifted model with noise is given in (4.8), where both $\underline{P}$ and $\underline{N}$ are MIMO but LTI, and the noise $\underline{e}$ satisfies the same standard assumptions as e does. One might think that identifying a state-space model for $\underline{P}$ based on data $\underline{u}$ and $\underline{y}$ is straightforward by applying subspace identification techniques; but this is *not* the case due to causality constraints on the D-matrix in state-space models of $\underline{P}(z)$. Because of this

complication, the subspace algorithms need to be re-derived from the least-square optimization, which is the focus of this section.

Suppose the continuous-time process involves a time delay of $d_c T$. From our discussion in the preceding section, $d_c T$ can be estimated within accuracy T. With this information, by shifting the multirate input and output data properly (as discussed towards the end of Section 2.4), we can assume without loss of generality that $0 < d_c \leq 1$ in this section.

Let $\underline{P}$ have the following state-space model

$$\underline{P}(z) = \left[\begin{array}{c|c} \underline{A} & \underline{B} \\ \hline \underline{C} & \underline{D} \end{array}\right]. \tag{4.15}$$

The $m \times n$ matrix $\underline{D}$ relates $\underline{u}(0)$ to $\underline{y}(0)$ as follows:

$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(n_i - 1) \end{bmatrix} = \begin{bmatrix} D_{00} & D_{01} & \cdots & D_{0,n-1} \\ D_{10} & D_{11} & \cdots & D_{1,n-1} \\ \vdots & \vdots & & \vdots \\ D_{m-1,0} & D_{m-1,1} & \cdots & D_{m-1,n-1} \end{bmatrix} \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(n-1) \end{bmatrix}. \tag{4.16}$$

Note that D_{ij} takes the input $u(j)$ at time $t = jmT$ to the output $y(i)$ which occurs at time $t = inT$. Because of the time delay of $d_c T$ in the system, causality requires that $D_{ij} = 0$ if $jmT + d_c T > inT$. Since $0 < d_c \leq 1$, we can write the causality constraint as follows:

$$D_{ij} = 0, \quad \text{whenever} \quad jm \geq in. \tag{4.17}$$

For example, if $m = 2$ and $n = 3$, the structure of $\underline{D}$ is given by

$$\underline{D} = \begin{bmatrix} 0 & 0 & 0 \\ D_{10} & D_{11} & 0 \end{bmatrix}.$$

In general, the causality condition implies that $\underline{D}$ must be block lower triangular: Given $0 < i \leq m-1$, define J_i as the largest j satisfying $jm < in$; it follows that $D_{ij} \neq 0$ for $j \leq J_i$, otherwise $D_{ij} = 0$. Defining

$$D_i = \begin{bmatrix} D_{i,0} & D_{i,1} & \cdots & D_{i,J_i} \end{bmatrix}, \quad i = 1, 2, \cdots, m-1, \tag{4.18}$$

we get that a causal $\underline{D}$ takes the following form:

$$\underline{D} = \begin{bmatrix} 0 & 0 \\ D_1 & 0 \\ \vdots & \vdots \\ D_{m-1} & 0 \end{bmatrix}. \tag{4.19}$$

Note that the first row is zero because of the time delay; the zero blocks in (4.19) may have different sizes. This structure of zero blocks needs to be enforced in the

identification process to guarantee causality. Next we discuss how to modify the subspace identification algorithms to incorporate this constraint on $\underline{D}$.

Let us start with the lifted state-space model

$$\begin{aligned} \underline{x}(k+1) &= \underline{A}\,\underline{x}(k) + \underline{B}\,\underline{u}(k) \\ \underline{y}(k) &= \underline{C}\,\underline{x}(k) + \underline{D}\,\underline{u}(k) + \underline{v}(k), \end{aligned}$$

where $\underline{v}$ is the output from the lifted noise model $\underline{N}$. Suppose we have N (lifted) data points. We first estimate the state vector $\underline{x}(k)$ using the CVA or N4SID algorithm; the causality constraint on $\underline{D}$ does not affect the state estimation. Next, we solve the least-square optimization problem to find a state-space model:

$$\min_{\underline{A},\underline{B},\underline{C},\underline{D}} \frac{1}{N}\sum_{k=0}^{N-1} \left\{ \begin{bmatrix} \underline{x}(k+1) \\ \underline{y}(k) \end{bmatrix} - \begin{bmatrix} \underline{A} & \underline{B} \\ \underline{C} & \underline{D} \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \underline{u}(k) \end{bmatrix} \right\}^T \left\{ \begin{bmatrix} \underline{x}(k+1) \\ \underline{y}(k) \end{bmatrix} - \begin{bmatrix} \underline{A} & \underline{B} \\ \underline{C} & \underline{D} \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \underline{u}(k) \end{bmatrix} \right\}.$$

In this optimization process, we now incorporate the causality condition in $\underline{D}$. Corresponding to (4.19), we partition $\underline{C}$, $\underline{u}$ and $\underline{y}$:

$$\underline{C} = \begin{bmatrix} C_0 \\ C_1 \\ \vdots \\ C_{m-1} \end{bmatrix}, \quad \underline{u} = \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{n-1} \end{bmatrix}, \quad \underline{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{m-1} \end{bmatrix}. \tag{4.20}$$

From here we define partial vectors for $\underline{u}$ according to the structures in (4.18):

$$\underline{u}_{J_i} = \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{J_i} \end{bmatrix}, \quad i = 1, 2, \cdots, m-1.$$

The least-square problem is equivalent to minimizing over $\{\underline{A}, \underline{B}, C_i, D_i\}$ the quantity

$$\begin{aligned} \frac{1}{N}\sum_{k=0}^{N-1} \Bigg\{ & \left(\underline{x}(k+1) - \begin{bmatrix} \underline{A} & \underline{B} \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \underline{u}(k) \end{bmatrix} \right)^T \left(\underline{x}(k+1) - \begin{bmatrix} \underline{A} & \underline{B} \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \underline{u}(k) \end{bmatrix} \right) \\ & + [y_0(k) - C_0\underline{x}(k)]^T [y_0(k) - C_0\underline{x}(k)] \\ & + \sum_{i=1}^{m-1} \left(y_i(k) - \begin{bmatrix} C_i & D_i \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \underline{u}_{J_i}(k) \end{bmatrix} \right)^T \left(y_i(k) - \begin{bmatrix} C_i & D_i \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \underline{u}_{J_i}(k) \end{bmatrix} \right) \Bigg\} \end{aligned}$$

The least-square solutions for $\underline{A}$, $\underline{B}$, C_0 and C_i, D_i $(i = 1, 2, \cdots, m-1)$ can be

computed by setting all the partial derivatives to zero:

$$\frac{1}{N}\sum_{k=0}^{N-1}\left(\underline{x}(k+1)-\left[\begin{array}{cc}\underline{A} & \underline{B}\end{array}\right]\left[\begin{array}{c}\underline{x}(k)\\ \underline{u}(k)\end{array}\right]\right)\left[\begin{array}{c}\underline{x}(k)\\ \underline{u}(k)\end{array}\right]^T=0,$$

$$\frac{1}{N}\sum_{k=0}^{N-1}\left[y_0(k)-C_0\underline{x}(k)\right]\underline{x}(k)^T=0,$$

$$\frac{1}{N}\sum_{k=0}^{N-1}\left(y_i(k)-\left[\begin{array}{cc}C_i & D_i\end{array}\right]\left[\begin{array}{c}\underline{x}(k)\\ \underline{u}_{J_i}(k)\end{array}\right]\right)\left[\begin{array}{c}\underline{x}(k)\\ \underline{u}_{J_i}(k)\end{array}\right]^T=0,\quad i=1,2,\cdots,m-1.$$

Solving the above equations, we find the optimal solution: Defining the square matrices

$$\begin{aligned}\Phi &= \frac{1}{N}\sum_{k=0}^{N-1}\left[\begin{array}{c}\underline{x}(k)\\ \underline{u}(k)\end{array}\right]\left[\begin{array}{c}\underline{x}(k)\\ \underline{u}(k)\end{array}\right]^T,\\ \Phi_0 &= \frac{1}{N}\sum_{k=0}^{N-1}\underline{x}(k)\underline{x}(k)^T,\\ \Phi_i &= \frac{1}{N}\sum_{k=0}^{N-1}\left[\begin{array}{c}\underline{x}(k)\\ \underline{u}_{J_i}(k)\end{array}\right]\left[\begin{array}{c}\underline{x}(k)\\ \underline{u}_{J_i}(k)\end{array}\right]^T,\quad i=1,2,\cdots,m-1,\end{aligned}$$

(note that under proper excitation conditions, these matrices should be nonsingular), we have

$$\begin{aligned}\left[\begin{array}{cc}\underline{A} & \underline{B}\end{array}\right] &= \left[\begin{array}{cc}\frac{1}{N}\sum_{k=0}^{N-1}\underline{x}(k+1)\underline{x}(k)^T & \frac{1}{N}\sum_{k=0}^{N-1}\underline{x}(k+1)\underline{u}(k)^T\end{array}\right]\Phi^{-1},\\ C_0 &= \left[\frac{1}{N}\sum_{k=0}^{N-1}\underline{y}_0(k)\underline{x}(k)^T\right]\Phi_0^{-1},\\ \left[\begin{array}{cc}C_i & D_i\end{array}\right] &= \left[\begin{array}{cc}\frac{1}{N}\sum_{k=0}^{N-1}y_i(k)\underline{x}(k)^T & \frac{1}{N}\sum_{k=0}^{N-1}y_i(k)\underline{u}_{J_i}(k)^T\end{array}\right]\Phi_i^{-1},\quad i=1,\cdots,m-1.\end{aligned}$$

We remark that this state-space model computed is the optimal causal one in the least-square sense.

4.6 Computing Fast Single-Rate Models

In the preceding section, we proposed a modified subspace identification algorithm for estimating a state-space model of the lifted multirate process; the model obtained is guaranteed to satisfy the causality constraint. In this section, we go one step further and show how to compute a single-rate model with sampling period mT (the control period) based on the lifted model. Such a single-rate model operates at the faster rate ($m < n$); as a special case, if $m = 1$, working at a rate which is n times the output sampling rate, this model can be used to estimate the missing output samples for inferential control applications [47, 27].

4.6.1 Controllability Approach

We start with a state-space model for $\underline{P}$ in (4.15). Knowing the matrices $(\underline{A}, \underline{B}, \underline{C}, \underline{D})$, we would like to compute the state-space model for $P_{mT} = S_{mT} P H_{mT}$; specifically, from (4.5), we would like to compute the matrices (A_{mT}, B_{mT}, C, D). According to lifting, partition $\underline{C}$ and $\underline{D}$ as in (4.20) and (4.16), respectively, and $\underline{B}$ as follows:

$$\underline{B} = \begin{bmatrix} B_0 & B_1 & \cdots & B_{n-1} \end{bmatrix}. \tag{4.21}$$

From Proposition 1, B_{mT}, C and D can be read off directly:

$$B_{mT} = B_{n-1} \quad C = C_0 \quad D = D_{00}.$$

Thus the question reduces to how to compute A_{mT} from the given data. We will discuss two methods for doing this.

The first approach to compute A_{mT} is based on the hypothesis that (A_{mT}, B_{mT}) is controllable. Given $\underline{A}$ and $\underline{B}$ in (4.21), Proposition 1 implies that

$$A_{mT}^n = \underline{A}, \quad B_{mT} = B_{n-1}, \quad A_{mT} B_{mT} = B_{n-2}, \cdots, A_{mT}^{n-1} B_{mT} = B_0.$$

Multiplying $\underline{A}$ to B_j, $\underline{A}^2$ to B_j, and so on, we can get $A_{mT}^k B_{mT}$ for any $k \geq 0$. Thus we can form the following two matrices

$$\begin{aligned} \Phi_c &= \begin{bmatrix} B_{mT} & A_{mT} B_{mT} & \cdots & A_{mT}^{p-1} B_{mT} \end{bmatrix}, \\ \Phi &= \begin{bmatrix} A_{mT} B_{mT} & A_{mT}^2 B_{mT} & \cdots & A_{mT}^p B_{mT} \end{bmatrix}, \end{aligned}$$

where p is the dimension of A_{mT}. The matrix Φ_c is the controllability matrix for the pair (A_{mT}, B_{mT}). Note that

$$\Phi = A_{mT} \Phi_c.$$

The controllability assumption implies that Φ_c has full row rank; therefore A_{mT} can be uniquely computed by

$$A_{mT} = \Phi \Phi_c^T (\Phi_c \Phi_c^T)^{-1}.$$

4.6.2 Eigenvalue Approach

The second approach to compute A_{mT} is based on an eigen-problem for $\underline{A}$, for which we assume $\underline{A}$ has distinct eigenvalues. The steps involved are summarized below:

1. Compute the eigenvalues and eigenvectors for $\underline{A}$ to get

$$V^{-1} \underline{A} V = \text{diag}\{\underline{\lambda}_0, \cdots, \underline{\lambda}_{p-1}\},$$

where columns in V are the eigenvectors.

2. Pre-multiply B_{n-1} and B_{n-2}, the last two columns in $\underline{B}$, by V^{-1} to get two column vectors

$$V^{-1}B_{n-1} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{p-1} \end{bmatrix}, \quad V^{-1}B_{n-2} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix}. \tag{4.22}$$

Let the eigenvalues of A_{mT} be λ_i, $i = 0, 1, \cdots, p-1$. The matrices A_{mT} and $\underline{A}$ share the same eigenvector matrix:

$$V^{-1}A_{mT}V = \text{diag}\{\lambda_0, \cdots, \lambda_{p-1}\}.$$

Since $B_{n-2} = A_{mT}B_{n-1}$ (Proposition 1), we can compute the eigenvalues λ_i from the two vectors in (4.22):

$$\lambda_i = b_i/a_i, \quad i = 0, 1. \cdots, p-1.$$

3. Thus we get A_{mT}

$$A_{mT} = V\text{diag}\{\lambda_0, \cdots, \lambda_{p-1}\}V^{-1}.$$

4.6.3 Matrix Roots Approach

Note that if a unique A_{mT} can be obtained from A_{mT}^n, then the single-rate model with sampling interval mT is available. But we know that a unique A_{mT} from A_{mT}^n can not always be obtained directly by taking the n^{th} root of A_{mT}^n. Assume that α is a pole of the continuous process $P_c(s)$. The corresponding pole of the system discretized with interval mT is

$$\sigma_i = e^{mT\alpha}$$

and the corresponding pole of a system discretized with interval nmT is

$$\rho_i = e^{nmT\alpha}$$

ρ_i can be expressed as:

$$\rho_i = re^{i\theta}, \qquad -\pi \leq \theta \leq \pi$$

where r is the magnitude of ρ_i and θ is the main angle of ρ_i. There are a total of n roots of ρ_i:

$$\sqrt[n]{r}e^{i(2k\pi+\theta)/n}, k = 0, 1, ..., n-1$$

Assume that

$$\alpha = \gamma + i\eta$$

then

$$\begin{aligned} \rho_i &= e^{nmT\alpha} = e^{nmT\gamma}e^{inmT\eta} \\ \sigma_i &= e^{mT\alpha} = e^{mT\gamma}e^{imT\eta} \\ nmT\eta &= 2J\pi + \theta, \ -\pi \leq \theta \leq \pi \end{aligned}$$

where J is an integer ≥ 0. α is unknown and J is unknown, so a unique σ_i can not be computed from ρ_i. σ_i and ρ_i are eigenvalues of matrix A_{mT} and A^n_{mT} respectively, this means that matrix A_{mT} can not be computed uniquely from matrix A^n_{mT}. If we know that J is zero, then matrix A_{mT} can be computed uniquely from matrix A^n_{mT}.

Proposition 3 *A fast-sample model with interval mT can be computed from a lifted model with sampling interval nmT directly if the following requirement is satisfied*

$$|nmT\eta| \leq \pi$$

where η is the largest imaginary part of the poles of the continuous process.

Proof First we assume that the matrix A_{mT} is diagonalizable, then

$$A_{mT} = VSV^{-1}, \ \ A^n_{mT} = VS^nV^{-1}$$

The eigenvector matrix V of the matrix A_{mT} can then be computed from the similarity decomposition of matrix A^n_{mT}. Because we also assume that $|nmT\eta| \leq \pi$, then S can be computed from S^n uniquely. This means that matrix A_{mT} can be computed uniquely from matrix A^n_{mT}.

Now we assume that the matrix A_{mT} is not diagonalizable, this means that some of the eigenvalues of matrix A_{mT} are the same and the algebraic multiplicity of the multiple eigenvalue is larger than the geometric multiplicity of the multiple eigenvalue [26]. For simplicity let A_{mT} be a matrix with l repeated eigenvalues at λ. We know that matrix A_{mT} can be expressed as

$$\begin{aligned} A_{mT} &= C_1^{-1}J_1C_1 \\ J_1 &= \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}_{l\times l} \end{aligned}$$

where C_1 is an $l \times l$ invertible matrix and J_1 is the Jordan canonical form. A^n_{mT} can be expressed as,

$$\begin{aligned} A^n_{mT} &= \overbrace{A_{mT} \times A_{mT} \times \cdots \times A_{mT}}^{n} \\ &= C_1^{-1} J_1^n C_1 \\ &= C_1^{-1} J_2 C_1 \end{aligned}$$

where

$$J_2 = \begin{bmatrix} \lambda^n & (n+1)\lambda^{n-1} & n\lambda^{n-2} & \cdots & 4\lambda^2 & 3\lambda \\ 0 & \lambda^n & (n+1)\lambda^{n-1} & \cdots & 5\lambda^3 & 4\lambda^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda^n & (n+1)\lambda^{n-1} \\ 0 & 0 & 0 & \cdots & 0 & \lambda^n \end{bmatrix}$$

And A^n_{mT} can also be decomposed as:

$$A^n_{mT} = C_2^{-1} \begin{bmatrix} \lambda^n & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda^n & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda^n & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda^n \end{bmatrix} C_2$$

and matrix J_2 can be decomposed in the Jordan form as:

$$\begin{aligned} J_2 &= C_3^{-1} J_3 C_3 \\ J_3 &= \begin{bmatrix} \lambda^n & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda^n & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda^n & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda^n \end{bmatrix} \end{aligned}$$

Therefore,

$$A^n_{mT} = C_1^{-1} C_3^{-1} \begin{bmatrix} \lambda^n & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda^n & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda^n & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda^n \end{bmatrix} C_3 C_1$$

This means:

$$C_2 = C_3 C_1$$

We know matrix A^n_{mT}, so we can compute matrices C_2 and J_3. From the assumption, $|nmT\eta| \leq \pi$, it is obvious that λ can be computed from λ^n uniquely, J_2 can then be

formed and matrix C_3 can be computed from matrix J_2. After matrices C_2 and C_3 are known, matrix C_1 can be computed as follows,

$$C_1 = C_3^{-1} C_2$$

Once matrices C_1 and J_1 are obtained matrix A_{mT} is known.

If the condition $|nmT\eta| \leq \pi$ is satisfied, a unique A_{mT} can then be obtained from A_{mT}^n directly; and the input and the output matrices, B_{mT}, C and D, respectively, can be easily obtained from the appropriate partitions of the estimated lifted model $\underline{P}$, hence the single-rate model with sampling interval mT can be computed from the estimated lifted model $\underline{P}$. ■

The condition for the controllability method is relatively easy to satisfy; but when the condition number of the controllability matrix of the fast-sampled model is large, the estimation error of the controllability approach will also be large. The conditions for the eigenvalue approach and the matrix roots approach may not be always satisfied. Usually the matrix roots approach can give the best estimation.

4.7 Procedure Summary

In this section we summarize what we have studied in the preceding sections for identifying a fast-rate state-space model from the multirate data, when the continuous-time process is time-delayed.

Step 1 Estimate the time delay $d_c T$ of the continuous-time process based on multirate input-output data (Section 4); the error of this estimation can be as small as T, assuming m and n are co-prime.

Step 2 Suppose d_c is estimated to be within the interval $(d_d, d_d + 1]$, where d_d is a nonnegative integer. There always exist integers k_1 and k_2 such that

$$d_d = k_1 m + k_2 n;$$

then we shift the input data to the left by k_1 samples and the output data to the right by k_2 samples, so that the shifted input-output data gives rise to a new system with time delay no larger than T.

Step 3 Identify a lifted state-space model based on the shifted multirate data using the modified subspace identification algorithm that we proposed in Section 5; this guarantees the model satisfies the causality constraint, and is both controllable and observable under a mild condition on sampling (Section 3).

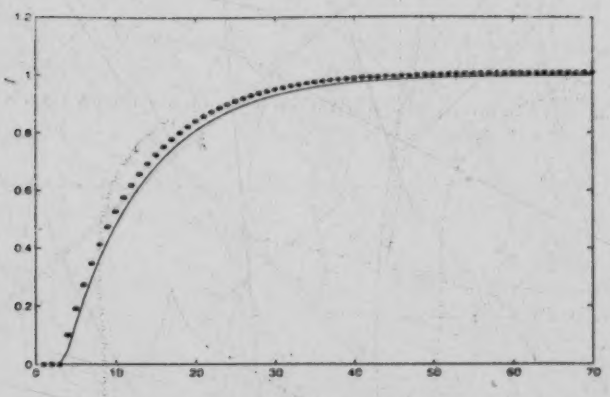


Figure 4.5: *Step responses of the actual (solid) and estimated (star) fast-rate models*

Step 4 Extract a fast single-rate model with sampling period mT from the lifted model, using either the controllability or eigenvalue approach (Section 6).

Step 5 Incorporate the time delay estimated in Step 1 to the fast-rate model.

4.8 Simulation Example

An illustrative example is given in this section; this example is a simple SISO system and the study is based on computer simulation.

Example 1. In view of Figure 4.1, take the process and noise models to be

$$P_c(s) = \frac{e^{-5s}}{20s+1}, \quad N(s) = \frac{1}{10s+1}.$$

The control period is 2 s; and the sampling period is 3 s ($m = 2$, $n = 3$, $T = 1$). Thus the lifted multirate system has 3 inputs and 2 outputs.

First, we generate a vector white noise sequence of length 3000 as the lifted input, measure the output with noise (the signal-to-noise ratio is 3:1); and then estimate the time delay. Next, we generate a (vector) lifted low-frequency excitation sequence of length 3000, again measure the output with signal-to-noise ratio 3:1. Shifting the input-output data, we estimate a lifted model and then a fast-rate model. Finally, we incorporate the time delay into the fast-rate model. Figure 4.5 compares step responses for the actual and estimated models with sampling period 2 s: Given the signal-to-noise ratio and the input excitation conditions, the estimation is fairly accurate.

4.9 Conclusion

In this chapter we have shown by a theoretical analysis and a simulation example, that a fast single-rate model in which the sampling interval is the same as the control interval of the multirate system can be estimated from multirate data collected from the general class of SISO multirate systems. The most general case of a SISO multirate system is one in which the ratio between the output sampling interval and the control interval is a rational number. The proposed methods for estimation have been successfully applied and evaluated on a simulated system. The results in this chapter differ from those by Verhaegen and Yu [82] in that we tackle the causality constraint explicitly in the optimization, analyze the observability of the lifted model and identify not only a lifted model for the multirate process but also a fast single-rate model, based on *multirate* input-output data.

Chapter 5

Subspace-based Identification of MIMO Multirate Systems

5.1 Introduction

In the chemical process industry, most processes are multi-input and multi-output (MIMO) systems. Strong interaction in the MIMO systems makes it difficult to control them with traditional single loop controllers such as PID ones. Model-based predictive controllers can handle the interaction, and that is one of the reasons why MPC has been widely accepted in the chemical process industry.

It is well known that the state space framework for single-rate SISO systems can be extended naturally to represent single-rate MIMO systems, hence subspace-based identification algorithms for single-rate SISO systems can be extended to single-rate MIMO systems directly. Researchers have shown that in the identification of MIMO systems, there are numerically ill-conditioned mathematical problems in the polynomial identification framework. However there are no such problems in the subspace-based identification framework.

In Chapter 4 we discussed the identification of SISO multirate systems. In this chapter we will discuss the identification of MIMO multirate systems. There are many classes of MIMO multirate systems; and we will only focus on one special class of MIMO multirate systems where all the control intervals are uniform, the output sampling intervals may be different, but are integer multiples of the control interval.

5.2 Identification of MIMO Multirate Systems

Consider the multi-input and multi-output multirate sampled-data system shown in Figure 5.1. Here, P_c is a continuous-time process with additive disturbances; the disturbances are generated by the continuous-time model N with white noise

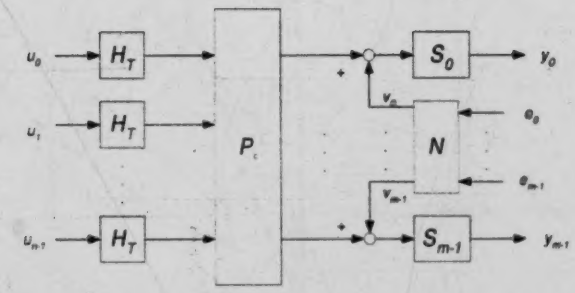


Figure 5.1: *Block diagram of a MIMO multirate sampled-data system*

inputs $e_0, \cdots, e_{m-1}$; the outputs are sampled by m samplers S_i with periods J_iT, $i = 0, 1, ..., m-1$, yielding the sampled outputs $y_i(k)$, $i = 0, 1, ..., m-1$; the inputs to P_c are generated by zero-order holds with period T processing the input sequences $u_i(k), i = 0, 1, ..., n-1$. We assume that J_i is an integer (≥ 1), for all $i = 0, 1, ..., m-1$.

Both $u_i(k)$ and $y_j(k)$ are discrete-time signals defined on the time set $\mathcal{Z}_+ := \{0, 1, 2, \cdots\}$; but their underlying periods are T and J_jT, respectively. Define

$$u = \begin{bmatrix} u_0 \\ \vdots \\ u_{n-1} \end{bmatrix}, \quad y = \begin{bmatrix} y_0 \\ \vdots \\ y_{m-1} \end{bmatrix}.$$

Assume that P_c is linear time-invariant (LTI). The discrete-time system from u to y (with $e_k = 0, k = 0, ..., m-1$), is linear and periodically time-varying. In order to get a tractable model for this multirate system, we use the lifting technique. For the multirate system from u to y in Figure 5.1, we lift the inputs and outputs to get a single-rate system with underlying period JT, where J is the least common multiple of $J_0, J_1, ..., J_{m-1}$; and hence we need to lift u_i by L_J, and y_j by L_{k_j} to arrive at Figure 5.2, where $k_j = \frac{J}{J_j}$. Let $S_{J,T}$ represent a sampler with interval J_iT, so $S_i = S_{J,T}$. For simplicity we assume that $n \geq m$. Define

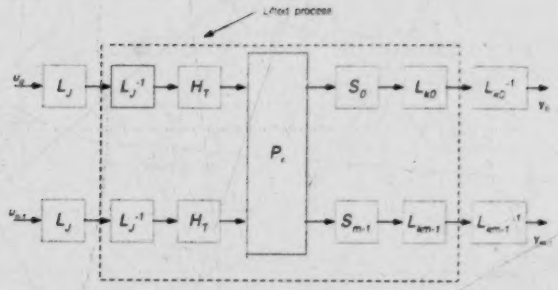


Figure 5.2: *The lifted MIMO multirate system*

$$\underline{u} = \begin{bmatrix} L_J(u_0) \\ \vdots \\ L_J(u_{n-1}) \end{bmatrix}, \quad \underline{y} = \begin{bmatrix} L_{k_0}(y_0) \\ \vdots \\ L_{k_{m-1}}(y_{m-1}) \end{bmatrix}.$$

The lifted multirate system $\underline{P}$ takes $\underline{u}$ into $\underline{y}$ and is defined via

$$\begin{aligned} \underline{P} &= \begin{bmatrix} L_{k_0} & & \\ & \ddots & \\ & & L_{k_{m-1}} \end{bmatrix} \begin{bmatrix} S_0 & & \\ & \ddots & \\ & & S_{m-1} \end{bmatrix} P_c \begin{bmatrix} H_T & & \\ & \ddots & \\ & & H_T \end{bmatrix} \begin{bmatrix} L_J^{-1} & & \\ & \ddots & \\ & & L_J^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \underline{P}_{00} & \cdots & \underline{P}_{0,n-1} \\ \underline{P}_{10} & \cdots & \underline{P}_{1,n-1} \\ \vdots & \ddots & \vdots \\ \underline{P}_{m-1,0} & \cdots & \underline{P}_{m-1,n-1} \end{bmatrix}. \end{aligned} \tag{5.1}$$

5.2.1 Lifted MIMO Multirate Systems

The $\underline{P}$ defined in equation (5.1) can be shown to be LTI. In order to derive a state-space model for $\underline{P}$, we discretize P_c via the zero-order hold with the sampling period T to get $P_T := S_T P_c H_T$, S_T and H_T being the sampler and zero-order hold with period T. P_T can be expressed as

$$P_T = \begin{bmatrix} P_{0,0} & P_{0,1} & \cdots & P_{0,n-1} \\ P_{1,0} & P_{1,1} & \cdots & P_{1,n-1} \\ \vdots & \vdots & \cdots & \vdots \\ P_{m-1,0} & P_{m-1,1} & \cdots & P_{m-1,n-1} \end{bmatrix}. \tag{5.2}$$

and can be represented by the state-space representation in the (A, B, C, D) form:

$$\left[\begin{array}{c|cccc} A & B_0 & B_1 & \cdots & B_{n-1} \\ \hline C_0 & D_{0,0} & D_{0,1} & \cdots & D_{0,n-1} \\ C_1 & D_{1,0} & D_{1,1} & \cdots & D_{1,n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{m-1} & D_{m-1,0} & D_{m-1,1} & \cdots & D_{m-1,n-1} \end{array}\right]. \tag{5.3}$$

Proposition 4 *A state-space model for the lifted multirate system* $\underline{P}$ *is given by*

$$\underline{P} = \left[\begin{array}{ccc} \underline{P}_{00} & \cdots & \underline{P}_{0,n-1} \\ \underline{P}_{10} & \cdots & \underline{P}_{1,n-1} \\ \vdots & \ddots & \vdots \\ \underline{P}_{m-1,0} & \cdots & \underline{P}_{m-1,n-1} \end{array}\right], \tag{5.4}$$

where

$$\underline{P}_{ij} = \left[\begin{array}{c|ccc} A^J & A^{J-1}B_i & A^{J-2}B_i & \cdots & B_i \\ \hline C_j & D_{i,j} & 0 & & 0 \\ C_jA^{J_i} & C_jA^{J_i-1}B_i & C_jA^{J_i-2}B_i & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ C_jA^{(k_i-1)J_i} & C_jA^{J-J_i-1}B_i & C_jA^{J-J_i-2}B_i & & 0 \end{array}\right].$$

Proof of Proposition 4 From (5.2) and the identities $S_{J,T} = S_{J,T}H_TS_T$, $i = 0, ..., m-1$, we can write

$$\begin{aligned} \underline{P} &= \left[\begin{array}{ccc} L_{k_0} & & \\ & \ddots & \\ & & L_{k_{m-1}} \end{array}\right] \left[\begin{array}{ccc} S_{J_0T}H_T & & \\ & \ddots & \\ & & S_{J_{m-1}T}H_T \end{array}\right] S_TP_cH_T \left[\begin{array}{ccc} L_J^{-1} & & \\ & \ddots & \\ & & L_J^{-1} \end{array}\right] \\ &= \left[\begin{array}{ccc} L_{k_0}S_{J_0T}H_T & & \\ & \ddots & \\ & & L_{k_{m-1}}S_{J_{m-1}T}H_T \end{array}\right] \left[\begin{array}{ccc} P_{0,0} & \cdots & P_{0,n-1} \\ \vdots & \vdots & \vdots \\ P_{m-1,0} & \cdots & P_{m-1,n-1} \end{array}\right] \left[\begin{array}{ccc} L_J^{-1} & & \\ & \ddots & \\ & & L_J^{-1} \end{array}\right]. \end{aligned}$$

The identity $L_J^{-1}L_J = I$ gives

$$\begin{aligned} \underline{P} &= \left[\begin{array}{ccc} L_{k_0}S_{J_0T}H_TL_J^{-1}L_JP_{00}L_J^{-1} & \cdots & L_{k_{m-1}}S_{J_{m-1}T}H_TL_J^{-1}L_JP_{0,n-1}L_J^{-1} \\ \vdots & \vdots & \vdots \\ L_{k_0}S_{J_0T}H_TL_J^{-1}L_JP_{m-1,0}L_J^{-1} & \cdots & L_{k_{m-1}}S_{J_{m-1}T}H_TL_J^{-1}L_JP_{m-1,n-1}L_J^{-1} \end{array}\right] \\ &= \left[\begin{array}{ccc} \cdots & \cdots & \cdots \\ \cdots & \underline{P}_{ij} & \cdots \\ \cdots & \cdots & \cdots \end{array}\right]; \end{aligned}$$

the last equality following from the definitions,

$$\begin{aligned} \underline{P}_{ij} &= \underline{S}_{ij}\underline{P}_{T,ij} \\ \underline{S}_{ij} &= L_{k_i}S_{J,T}H_TL_J^{-1} \end{aligned}$$

and

$$\underline{P}_{T,ij} = L_J P_{ij} L_J^{-1}.$$

It can be shown that $\underline{S}_{ij}$ is a static system with constant matrix representation:

$$\underline{S}_{ij} = \begin{bmatrix} \overbrace{I\ 0\ \dots\ 0}^{J_i} & \overbrace{0\ 0\ \dots\ 0}^{J_i} & \cdots & \overbrace{0\ 0\ \dots\ 0}^{J_i} \\ 0\ 0\ \dots\ 0 & I\ 0\ \dots\ 0 & \cdots & 0\ 0\ \dots\ 0 \\ \vdots\ \vdots\ \dots\ \vdots & \vdots\ \vdots\ \dots\ \vdots & \cdots & \vdots\ \vdots\ \dots\ \vdots \\ 0\ 0\ \dots\ 0 & 0\ 0\ \dots\ 0 & \cdots & I\ 0\ \dots\ 0 \end{bmatrix}_{k_i \times J}.$$

Now based on the state-space model of P_T in (5.3), a state-space model for $\underline{P}_{T,ij}$ is given by [42]:

$$\left[\begin{array}{c|cccc} A^J & A^{J-1}B_i & A^{J-2}B_i & \cdots & B_i \\ \hline C_j & D_{i,j} & 0 & \cdots & 0 \\ C_jA & C_jB_i & D_{i,j} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ C_jA^{J-1} & C_jA^{J-2}B_i & C_jA^{J-3}B_i & \cdots & D_{i,j} \end{array}\right]. \tag{5.5}$$

Pre-multiplying the transfer matrix in (5.5) by $\underline{S}_{ij}$ gives a state-space model for $\underline{P}_{ij}$:

$$\left[\begin{array}{c|cccc} A^J & A^{J-1}B_i & A^{J-2}B_i & \cdots & B_i \\ \hline C_j & D_{i,j} & 0 & & 0 \\ C_jA^{J_i} & C_jA^{J_i-1}B_i & C_jA^{J_i-2}B_i & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ C_jA^{(k_i-1)J_i} & C_jA^{J-J_i-1}B_i & C_jA^{J-J_i-2}B_i & & 0 \end{array}\right].$$

The state-space model for the lifted multirate system $\underline{P}$ can then be written as

$$\left[\begin{array}{c|cccccc} A^J & A^{J-1}B_0 & \dots & B_0 & \dots & B_{n-1} \\ \hline C_0 & D_{0,0} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ C_0A^{(k_0-1)J_0} & C_0A^{J-J_0-1}B_0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ C_{m-1} & D_{0,m-1} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ C_{m-1}A^{(k_{m-1}-1)J_{m-1}} & C_{m-1}A^{J-J_0-1}B_0 & \dots & 0 & \dots & 0 \end{array}\right].$$

■

After discussing lifting as applied to deterministic systems, we now study how to lift noise models. Here we assume that in Figure 5.1 e_i is the output of $e_i(k)$ processed

by a zero-order-hold with period J_iT, $i = 0, 1, ..., m-1$; so the noise process becomes a system with m distinct input sampling periods and m distinct output sampling periods; and we can lift the noise model by using the results in chapter 2. If we assume that the disturbance model N is a general transfer function matrix:

$$\begin{bmatrix} N_{0,0} & N_{0,1} & \cdots & N_{0,m-1} \\ \vdots & \vdots & \ddots & \vdots \\ N_{m-1,0} & N_{m-1,1} & \cdots & N_{m-1,m-1} \end{bmatrix},$$

then we need to discretize the noise model with sampling period T, and the lifting procedure becomes complicated since all the input periods are distinct and all the output sampling periods are distinct. For simplicity, we assume that N can be represented by a diagonal transfer function matrix:

$$\begin{bmatrix} N_0 & & \\ & \ddots & \\ & & N_{m-1} \end{bmatrix}. \tag{5.6}$$

This is not a restrictive assumption and most of the identification functions in MATALB such as ARX, ARMAX, BJ, ... make the same assumption.

In order to obtain the lifted noise model, based on the assumption that e_i is the output of $e_i(k)$ processed by a zero-order-hold with period J_iT, we first discretize the continuous-time noise model N_i in (5.6) with sampling period J_iT to get N_{i,J_iT} which converts the white noise term $e_i(k)$ into $v_i(k)$; assume a state-space model,

$$N_{i,J_iT}(z) = \left[\begin{array}{c|c} A_{i,J_i} & B_{i,J_i} \\ \hline C_{i,J_i} & D_{i,J_i} \end{array}\right],$$

$i = 0, 1, ..., m-1$. Since the deterministic system is lifted to interval JT, we lift N_{i,J_iT} by L_{k_i}. Defining $\underline{v_i} = L_{k_i} v_i$ and $\underline{e_i} = L_{k_i} e_i$ to obtain the lifted noise model, $\underline{N_i} := L_{k_i} N_{i,J_iT} L_{k_i}^{-1}$, such that it maps $\underline{e_i}$ to $\underline{v_i}$ and admits a state-space model:

$$\underline{N_i}(z) = \left[\begin{array}{c|ccc} A_{i,J_i}^{k_i} & A_{i,J_i}^{k_i-1} B_{i,J_i} & A_{i,J_i}^{k_i-2} B_{i,J_i} & \cdots & B_{i,J_i} \\ \hline C_{i,J_i} & D_{i,J_i} & 0 & \cdots & 0 \\ C_{i,J_i} A_{i,J_i} & C_{i,J_i} B_{i,J_i} & D_{i,J_i} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{i,J_i} A_{i,J_i}^{k_i-1} & C_{i,J_i} A_{i,J_i}^{k_i-2} B_{i,J_i} & C_{i,J_i} A_{i,J_i}^{k_i-3} B_{i,J_i} & \cdots & D_{i,J_i} \end{array}\right].$$

Define

$$\underline{e} = \begin{bmatrix} L_{k_0} e_0 \\ \vdots \\ L_{k_{m-1}} e_{m-1} \end{bmatrix}, \quad \underline{v} = \begin{bmatrix} L_{k_0} v_0 \\ \vdots \\ L_{k_{m-1}} v_{m-1} \end{bmatrix}.$$

The lifted noise model is

$$\underline{N} = \begin{bmatrix} L_{k_0} & & \\ & \ddots & \\ & & L_{k_{m-1}} \end{bmatrix} \begin{bmatrix} N_0 & & \\ & \ddots & \\ & & N_{m-1} \end{bmatrix} \begin{bmatrix} L_{k_0}^{-1} & & \\ & \ddots & \\ & & L_{k_{m-1}}^{-1} \end{bmatrix}$$
$$= \begin{bmatrix} L_{k_0} N_0 L_{k_0}^{-1} & & \\ & \ddots & \\ & & L_{k_{m-1}} N_{m-1} L_{k_{m-1}}^{-1} \end{bmatrix},$$

and can be obtained by combining the models of $\underline{N}_i, i = 0, 1, ..., m-1$. Thus we get the overall lifted model as follows:

$$\underline{y} = \underline{P}\,\underline{u} + \underline{N}\,\underline{e}. \tag{5.7}$$

Both $\underline{P}$ and $\underline{N}$ are LTI; and all statistical properties of $e_i(k)$ are preserved after lifting.

5.2.2 Extracting Fast Single-rate Models

In Chapter 4 we discuss the obervability of the lifted system in the presence of time delay, and show that the lifted state-space model derived from the fast-sampled state-space model will lose observability when the time delay of the process is larger than the control interval. The issues of observability in the presence of time delay are much more complex for MIMO systems. So for simplicity, in this chapter we assume that the time delay of the process is less than or equal to the control interval.

A method similar to the method discussed in Chapter 4 can be applied to overcome the causality constraint when we estimate the lifted model. Let $\underline{P}$ have the following state-space model

$$\underline{P}(z) = \left[\begin{array}{c|c} \underline{A} & \underline{B} \\ \hline \underline{C} & \underline{D} \end{array}\right].$$

After estimating the lifted model, we need to go one step further to compute the fast model with sampling interval T from the lifted model; specifically, from (5.4), we would like to compute matrices (A, B, C, D) in the following form:

$$\left[\begin{array}{c|ccc} A & B_0 & \cdots & B_{n-1} \\ \hline C_0 & D_{0,0} & \cdots & D_{0,n-1} \\ \vdots & \vdots & \vdots & \vdots \\ C_{m-1} & D_{m-1,0} & \cdots & D_{m-1,n-1} \end{array}\right].$$

Define $f = \sum_{i=0}^{m-1} k_i$; the lifted system has nJ inputs and f outputs. According to

the lifted inputs and outputs, partition $\underline{B}$, $\underline{C}$ and $\underline{D}$ as:

$$\underline{C} = \begin{bmatrix} \underline{C}_0 \\ \vdots \\ \underline{C}_{f-1} \end{bmatrix}, \quad \underline{B} = \begin{bmatrix} \underline{B}_0 & \cdots & \underline{B}_{nJ-1} \end{bmatrix}$$

$$\underline{D} = \begin{bmatrix} \underline{D}_{00} & \cdots & \underline{D}_{0,nJ-1} \\ \vdots & \ddots & \vdots \\ \underline{D}_{f-1,0} & \cdots & \underline{D}_{f-1,nJ-1} \end{bmatrix}.$$

From Proposition 4, matrices B, C and D can be read off directly as:

$$B_0 = \underline{B}_{J-1}, \; B_1 = \underline{B}_{2J-1}, \cdots, \; B_{n-1} = \underline{B}_{nJ-1}$$

$$C_0 = \underline{C}_0, \; C_1 = \underline{C}_{k_0}, \cdots, \; C_{m-1} = \underline{C}_{f-k_{m-1}}$$

$$D_{0,0} = \underline{D}_{0,0}, \; D_{01} = \underline{D}_{0,J}, \cdots, \; D_{m-1,0} = \underline{D}_{f-k_{m-1},0}, \cdots, \; D_{m-1,n-1} = \underline{D}_{f-k_{m-1},(n-1)J}.$$

Now the question that arises is how to compute matrix A from the lifted model. Clearly with the results in Chapter 4 we know that if the condition

$$|JT\eta| \leq \pi$$

(where η is the largest imaginary part of the poles of the continuous process P_c), is satisfied, we can compute matrix A directly from matrix $\underline{A}$; or if $\underline{A}$ has distinct eigenvalues, then we can also compute matrix A directly from matrix $\underline{A}$. In this section we will discuss only how to compute matrix A through the controllability approach: First we assume that the pair (A, B) is controllable. Given $\underline{A}$ and $\underline{B}$ in (5.4), Proposition 4 implies that

$$A^J = \underline{A}, \; B = \begin{bmatrix} \underline{B}_{J-1} & \cdots & \underline{B}_{nJ-1} \end{bmatrix}$$

$$AB = \begin{bmatrix} \underline{B}_{J-2} & \cdots & \underline{B}_{nJ-2} \end{bmatrix}, \cdots, \; A^{J-1}B = \begin{bmatrix} \underline{B}_0 & \cdots & \underline{B}_{(n-1)J} \end{bmatrix}.$$

Multiplying $\underline{A}$ to $A^{i-1}B$, $\underline{A}^2$ to $A^{i-1}B$, and so on, we can get A^kB for any $k \geq 0$. Thus we form the following two matrices,

$$\Phi_c = \begin{bmatrix} B & AB & \cdots & A^{p-1}B \end{bmatrix}$$

and

$$\Phi = \begin{bmatrix} AB & A^2B & \cdots & A^pB \end{bmatrix},$$

where p is the dimension of A. The matrix Φ_c is the controllability matrix for the pair (A, B). Note that

$$\Phi = A\Phi_c.$$

The controllability assumption implies that Φ_c has full row rank; therefore A can be uniquely computed by

$$A = \Phi\Phi_c^T(\Phi_c\Phi_c^T)^{-1}.$$

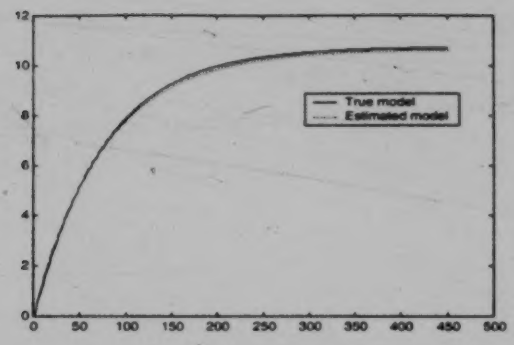


Figure 5.3: *Step responses of the true model (G11) and the estimated model*

5.3 Simulation and Experimental Examples

5.3.1 A Simulation Example

A 2-input and 2-output simulation sampled-data system is generated SIMULINK; the interval of holds is 1 unit of time; one of the output is sampled every unit of time, but the other is sampled every 5 units of time. The process model with sampling interval unit of time is

$$y(k) = \begin{bmatrix} \frac{0.1416z^{-1}}{1-0.9868z^{-1}} & \frac{-0.03z^{-1}}{1-0.97z^{-1}} \\ \frac{-0.0379z^{-1}}{1-0.9131z^{-1}} & \frac{0.0595z^{-1}}{1-0.9048z^{-1}} \end{bmatrix} u(k) + \begin{bmatrix} \frac{0.11z^{-1}}{1-0.97z^{-1}} & \frac{-0.03z^{-1}}{1-0.95z^{-1}} \\ \frac{-0.02z^{-1}}{1-0.9z^{-1}} & \frac{0.04z^{-1}}{1-0.88z^{-1}} \end{bmatrix} e(k).$$

A total of 10000 input-output data were collected with a random binary input sequence. The signal to noise ratio is 5. From this multirate data set we identified a fast-sampled single-rate model with a sampling interval of 1 unit of time, by applying the techniques developed in this chapter. The state-space model identified is of order 4, and the step responses of the true fast models and the estimated fast models are compared in Figures 5.3 to 5.6. The comparison shows that the estimated fast models are close to the true process models.

5.3.2 An Experimental Example

The MIMO process shown in Figure 5.7 is a computer-controlled experimental setup at the University of Alberta and is concerned with temperature and water level regulation in a continuously stirred tank heater.

In this process, the two manipulated variables are the positions of the cold water (u_1) and the steam (u_2) valves; the two measured variables are the water level (y_1)

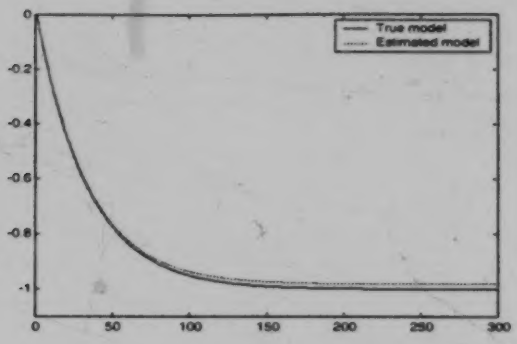


Figure 5.4: *Step responses of the true model (G12) and the estimated model*

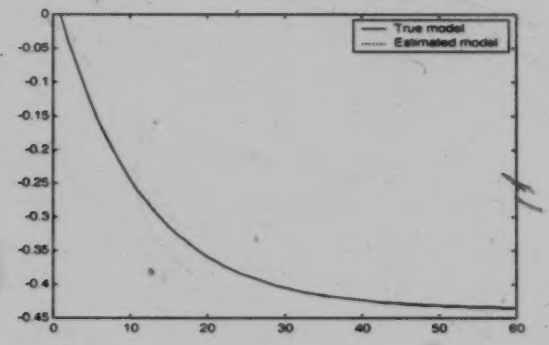


Figure 5.5: *Step responses of the true model (G21) and the estimated model*

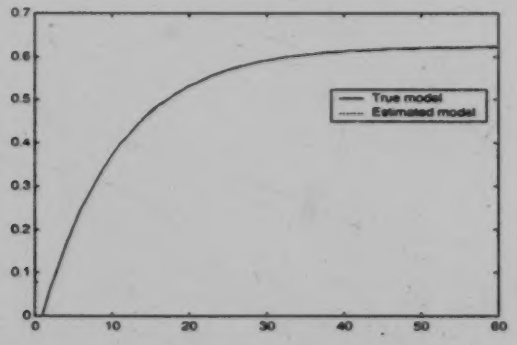


Figure 5.6: *Step responses of the true model (G22) and the estimated model*

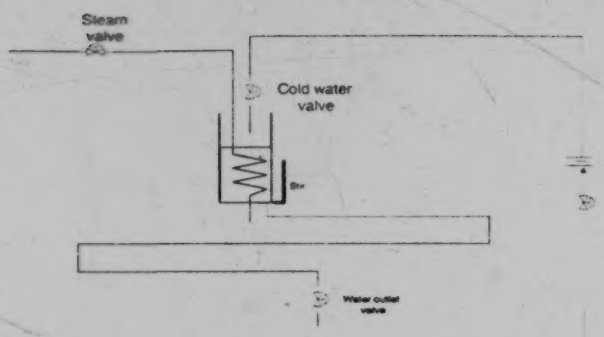


Figure 5.7: *Experimental setup of a stirred tank heater*

and water temperature (y_2) in the tank. For this study, the two valve positions are updated every 4 second, the water level and water temperature are sampled every 20 second and 4 second, respectively. Around the operating point with $y_1 = 12$ mA and $y_2 = 10$ mA, the inputs and outputs of the stirred tank heater are summarized below:

Symbol	Quantity	Range
u_1	cold water valve	$4 \text{ mA} \leq u_1 \leq 20 \text{ mA}$
u_2	steam valve	$4 \text{ mA} \leq u_2 \leq 20 \text{ mA}$
y_1	water level	$4 \text{ mA} \leq y_1 \leq 20 \text{ mA}$
y_2	water temperature	$4 \text{ mA} \leq y_2 \leq 20 \text{ mA}$

We use mA to quantify both y_1 and y_2 since there are simple linear relationships to translate these units to actual physical units. We collected the input-output data over 4 hours with a random binary input sequence. From this multirate data set we identified a fast-sampled single-rate model with a sampling interval of 4 second, by applying the techniques developed in this chapter. The state-space model identified is of order 4. Then we changed the sampling interval of the water level to 4 second, and collected the input-output data over 1 and half hours with a random binary input sequence. We used this fast-sampled data set to validate the estimated fast model. The model outputs and the actual process outputs (measured) are compared in Figures 5.8 and 5.9. It is clear from this validation test that the fast-rate model obtained from the multirate data captures the process dynamics very well.

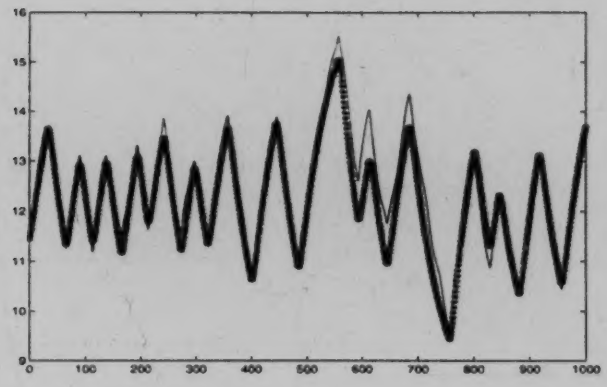


Figure 5.8: *Water level plots for the model (star) and the actual process (solid)*

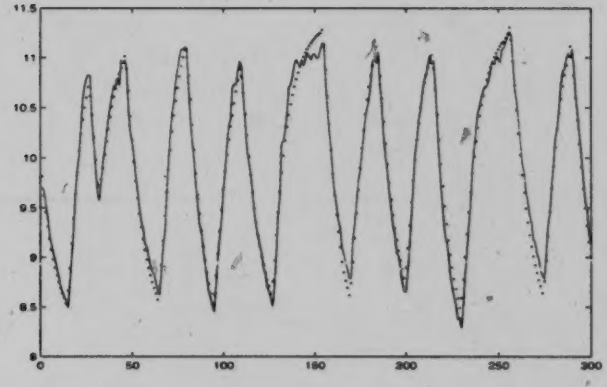


Figure 5.9: *Water temperature plots for the model (dot) and the actual process (solid)*

5.4 Conclusion

This chapter discusses a special case of MIMO multirate systems, where all the control intervals are uniform, all the output sampling intervals may be different, but integer multiples of the control interval. The analysis in this chapter shows that the lifted systems can be identified and single-rate models in which all the sampling intervals are the same as the control interval can be extracted from the lifted models.

Chapter 6

Multirate System Identification in the Polynomial Domain

6.1 Introduction

Identification of dynamic process models in the polynomial domain from measured input-output data is a well developed area; literature on this topic is extensive. Most of the research work in this area can be divided into two classes: prediction error methods (PEM) and instrumental variable (IV) methods. Prediction error methods have been widely accepted and practiced in industry. A major advantage of prediction error methods is that the method and its results can be interpreted in the frequency domain; and hence adds another dimension to the process insight.

Clearly the selection of a parameterized set of models is vital for the prediction error method. The simple ARX (AR refers to the auto-regressive part and X to the extra input) models have been widely studied for a long time [2, 60, 36]. The ARMAX models were introduced into system identification by Åström and Bohlin [3] and have been since then treated as simple models. Box and Jenkins developed the Box and Jenkins (BJ) models [6] which are more general than the ARMAX models. ARX models, ARMAX models and Output error (OE) models [39, 21] can be considered as the simplified forms of the most general family of BJ models [54]. The convergence and consistency for prediction-error estimates have been analyzed by Hannan [32], Ljung [52, 53] and Caines [8].

The frequency-domain expressions for the prediction-error criteria were first presented by Whittle in 1951 [85]; Whitte's work only dealt with the input-free case. Solo [70], Ljung and Glover [56] considered the case where there are extra inputs. More details on this topic can be found in [55].

The instrumental-variable (IV) methods were first introduced into statistics and

econometrics [65] and then applied to parameter estimation of dynamic systems [89, 86, 59]. Söderström and Stoica [69] analyzed the convergence and consistency aspects of the IV methods in a very comprehensive manner.

The methods mentioned are all single-rate based. For multirate systems, Lu and Fisher [57] considered a SISO multirate system where the output-sampling interval is an integer multiple of the control interval, and developed a method to estimate the relation between the future fast-sampled output and the measured inputs and outputs in the polynomial domain. Not much work has been done to estimate lifted models or fast-sampled models for multirate systems in the polynomial domain.

In Chapters 4 and 5 we proposed a two-step strategy in the state-space framework to estimate the lifted model from input-output data and then extract a fast-sampled model from the lifted model. We have shown that the lifted state-space model derived from the fast-sampled state-space model will lose observability when the time delay of the process is larger than the control interval. So in order to extract the fast-sampled model from the estimated lifted model when the time delay is large, we need to estimate the time delay first. But in practice, due to poor signal-to-noise ratio and other difficulties, it is not easy to estimate the time delay accurately. In order to overcome this disadvantage, we develop a two-step multirate identification method in the polynomial domain: identify lifted models from input-output measurements and then extract fast-sampled models. We will prove that the two-step multirate identification method in the polynomial domain does not have the time-delay-related observability problem. For simplicity we will only consider a special class of multirate systems where the ratio between the output sampling period and the control period is an integer.

This chapter is organized as follows. In Section 2 we present a multirate system identification scheme in the transfer function framework. In Section 3 we analyze the multirate identification algorithm in the frequency domain. Then we illustrate and validate the results of this chapter in Section 4 on an industrial case study. Finally concluding remarks are given in Section 5.

6.2 Multirate System Identification in the Transfer Function Domain

Consider the multirate sampled-data system in Figure 6.1. Here, P_c is a continuous-time process with additive noise; the noise is generated by a continuous-time model N with a white noise input e; the output of P_c is corrupted by that of N, and is sampled by a sampler S_{nT} with period nT, yielding the sampled output $y(k)$; the input to P_c

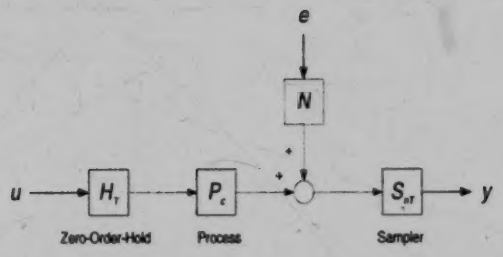


Figure 6.1: *Block diagram of a special SISO multirate sampled-data system*

is generated by a zero-order hold with period T processing the input sequence $u(k)$. Here we assume that $n(>1)$ is an integer which is a special but practical case of multirate systems. From now on, we will focus on this special case.

In chapter 4 we showed that if the single-rate model of the process P_c sampled with interval T is

$$D + C(zI - A)^{-1}B \triangleq \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right],$$

then the model of the lifted system $\underline{P}$ can be expressed as

$$\underline{D} + \underline{C}(zI - \underline{A})^{-1}\underline{B} = \left[\begin{array}{c|cccc} A^n & A^{n-1}B & A^{n-2}B & \cdots & B \\ \hline C & D & 0 & \cdots & 0 \end{array}\right]. \tag{6.1}$$

Discretize the continuous-time noise model N in Figure 6.1 with sampling period nT to get N_{nT}. Thus we get the overall lifted model as follows:

$$y(k) = \underline{P}(z^{-1})\,\underline{u}(k) + N_{nT}(z^{-1})e(k). \tag{6.2}$$

Both $\underline{P}$ and N_{nT} are LTI, so the identification algorithms in the polynomial domain can be applied to estimate the lifted model.

There is also a causality constraint when estimating the lifted transfer function model: Assume that the lifted model can be represented by a transfer function matrix:

$$\left[\begin{array}{cccc} P_0(z) & P_1(z) & \cdots & P_{n-1}(z) \end{array}\right].$$

Equation (6.1) means that transfer functions $P_1(z)$, ..., $P_{n-1}(z)$ should have at least first-order time delay. This causality constraint can easily be overcome by setting the time-delays of $P_1(z)$, ..., $P_{n-1}(z)$ to numbers larger than or equal to 1.

If the model of a lifted multirate system in Figure 6.1 can be estimated, then the natural question to ask is: can one determine the model of the fast-rate system from the estimated lifted model? The answer to this question is in the affirmative; this problem has been solved in the state-space domain [49]; and in this section we will

present a method for extracting a fast model in the transfer function domain: We start with a model for $\underline{P}$ in equation (6.1). Knowing the transfer function $\underline{P}(z)$, we would like to compute the model for $P_T = S_T P_c H_T$: specifically, from equation (6.1), we would like to compute the transfer function $P_T(z)$. $\underline{P}$ has n inputs, so $\underline{P}$ has n sub-systems:

$$\underline{P} = \begin{bmatrix} P_0 & P_1 & \cdots & P_{n-1} \end{bmatrix}.$$

Equation (6.1) means that

$$\begin{aligned} P_0(z) &= D + C(zI - A^n)^{-1}A^{n-1}B \\ P_1(z) &= C(zI - A^n)^{-1}A^{n-2}B \\ &\vdots \\ P_{n-1}(z) &= C(zI - A^n)^{-1}B. \end{aligned} \tag{6.3}$$

After estimating the transfer function of the lifted system, we can formulate a new transfer function, $\varphi(z)$, as follows

$$\begin{aligned} \varphi(z) &= P_0(z^n) + zP_1(z^n) + \cdots + z^{n-1}P_{n-1}(z^n) \\ &= C(z^nI - A^n)^{-1}[A^{n-1} + zA^{n-2} + \cdots + z^{n-1}I]B + D. \end{aligned} \tag{6.4}$$

Since

$$z^nI - A^n = (A^{n-1} + zA^{n-2} + \cdots + z^{n-1}I)(zI - A),$$

equation (6.4) can be reduced to

$$\varphi(z) = C(zI - A)^{-1}B + D.$$

Clearly, $\varphi(z)$ is exactly the single-rate model of the process sampled at the fast rate.

In Chapter 4 we concluded that when the time delay of the process is larger than the control interval, then the lifted model in equation (6.1) is not observable, and hence models of $P_0(z)$, ..., $P_{n-1}(z)$ will not have the structure in equation (6.3). As shown in Chapter 4, one of the solutions is to estimate the time-delay and then shift the input-output sequences according to the estimated time-delay, the lifted model in (6.1) of the new system is then observable. But in the polynomial domain this input-output shifting is unnecessary, and we can still calculate the fast-sampled model of the process from the models of the subsystems P_0, P_1, ..., P_{n-1}. This can be proved through the following arguments: Assume that transfer function $P_T(z)$ represents the discrete model of process P with sampling interval T; we know that $P_T(z)$ can be expressed as:

$$P_T(z) = p(0) + z^{-1}p(1) + z^{-2}p(2) + ...,$$

where z is the forward shift operator. Using the polyphase decomposition [74], we define

$$\begin{aligned}
\tilde{P}_0(z^n) &\triangleq p(0) + z^{-n}p(n) + z^{-2n}p(2n) + ... \\
\tilde{P}_1(z^n) &\triangleq p(1) + z^{-n}p(n+1) + z^{-2n}p(2n+1) + ... \\
&... \\
\tilde{P}_{n-1}(z^n) &\triangleq p(n-1) + z^{-n}p(2n-1) + z^{-2n}p(3n-1) + ...
\end{aligned}$$

Clearly, $\tilde{P}_1(z)$, $\tilde{P}_2(z)$, ..., and $\tilde{P}_{n-1}(z)$ are the models of systems P_0, P_1, ..., P_{n-1}, and

$$\begin{aligned}
\varphi(z) &= \tilde{P}_0(z^n) + z^{-1}\tilde{P}_1(z^n) + ... + z^{1-n}\tilde{P}_{n-1}(z^n) \\
&= P_T(z) \qquad (6.5)
\end{aligned}$$

Since the derivation of equation (6.5) is independent of the time-delay, so $\varphi(z)$ is the fast sampled model even when the time-delay of the process is larger than T.

6.3 Analysis in the Frequency Domain

Frequency domain analysis is of great importance, since it provides additional insight into the process dynamics. We will briefly analyze the proposed multirate identification in the frequency domain in this section.

Assume that a process model,

$$y(k) = \hat{\underline{P}}\underline{u}(k) + \hat{N}_{nT}e(k), \qquad (6.6)$$

is estimated from the multirate input-output data, where $\hat{\underline{P}}$ represents the estimated lifted deterministic model, and $\hat{N}$ represents the estimated disturbance model. Equation (6.6) can be used to predict the output. Let $\hat{y}(k|\theta)$ represent the predicted output, we have

$$\begin{aligned}
\hat{y}(k|\theta) &= \hat{\underline{P}}\underline{u}(k) + (\hat{N}_{nT} - 1)e \\
&= \hat{\underline{P}}\underline{u}(k) + (\hat{N}_{nT} - 1)\frac{1}{\hat{N}_{nT}}(y - \hat{\underline{P}}\underline{u}(k)) \\
&= y - \frac{1}{\hat{N}_{nT}}(y - \hat{\underline{P}}\underline{u}(k)), \qquad (6.7)
\end{aligned}$$

so the prediction error is

$$\hat{e}(k) = y(k) - \hat{y}(k|\theta) = \frac{1}{\hat{N}_{nT}}[y(k) - \hat{\underline{P}}\underline{u}(k)]. \qquad (6.8)$$

Substituting (6.2) and (6.7) into (6.8) gives

$$\begin{aligned}\hat{e}(k) &= \frac{1}{\hat{N}_{nT}}[\underline{P}\underline{u}(k) + N_{nT}e(k) - \underline{\hat{P}}\underline{u}(k)] \\ &= \frac{1}{\hat{N}_{nT}}[(\underline{P} - \underline{\hat{P}})\underline{u}(k) + N_{nT}e(k)] \\ &= \frac{1}{\hat{N}_{nT}}(\underline{P} - \underline{\hat{P}})\underline{u}(k) + \frac{N_{nT}}{\hat{N}_{nT}}e(k).\end{aligned}$$

Prediction error methods try to minimize the sum of the square of the prediction error,

$$\hat{\theta} = \arg_\theta \min \frac{1}{M}\sum_{k=1}^{M}(\hat{e}(k))^2. \tag{6.9}$$

Define

$$\begin{aligned}R_{\underline{u}}(\tau) &= \lim_{M\to\infty}\frac{1}{M}\sum_{t=1}^{M}\underline{u}(t)\underline{u}^T(t-\tau), \\ \phi_{\underline{u}} &= \sum_{\tau=-\infty}^{\infty}R_{\underline{u}}(\tau)e^{-j\omega\tau}.\end{aligned}$$

For simplicity we assume that $\underline{u}(k)$ is independent of $e(k)$. As the number of data points $M \to \infty$, we can apply Parseval's theorem to equation (6.9) and have

$$\hat{\theta} = \arg_\theta \min \int_{-\pi}^{\pi}\frac{1}{\left|\hat{N}_{nT}(e^{j\omega})\right|^2}\{[\underline{P}(e^{j\omega})-\underline{\hat{P}}(e^{j\omega})]\phi_{\underline{u}}[\underline{P}^T(e^{j\omega})-\underline{\hat{P}}^T(e^{j\omega})]+\left|N_{nT}(e^{j\omega})\right|^2\phi_e\}d\omega,$$

where ϕ_e is defined similarly. Since $\underline{u}$ is an $n \times 1$ vector, $\phi_{\underline{u}}$ is an $n \times n$ matrix. Singular value decomposition (SVD) of $\phi_{\underline{u}}$ yields

$$\phi_{\underline{u}} = USV^T,$$

where S is a diagonal matrix of the same dimension as $\phi_{\underline{u}}$, and with nonnegative diagonal elements in decreasing order; U and V are unitary matrices. $\phi_{\underline{u}}$ is a symmetric matrix which implies that U and V are identical, hence

$$\phi_{\underline{u}} = USU^T.$$

If

$$\underline{P}(e^{j\omega}) - \underline{\hat{P}}(e^{j\omega}) = [\ 0\ \cdots\ 0\ \ 1\ ]\,U^{-1},$$

then

$$\begin{aligned}[\underline{P}(e^{j\omega}) - \underline{\hat{P}}(e^{j\omega})]\phi_{\underline{u}}[\underline{P}^T(e^{j\omega}) - \underline{\hat{P}}^T(e^{j\omega})] &= [\ 0\ \cdots\ 0\ \ 1\ ]\,S\begin{bmatrix}0\\ \vdots\\ 0\\ 1\end{bmatrix} \\ &= \sigma_n\end{aligned}$$

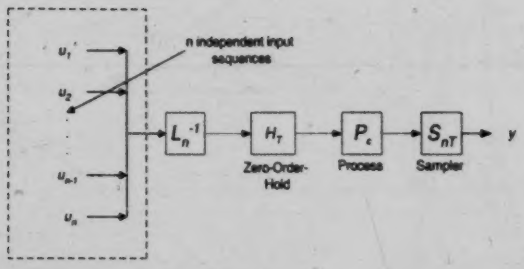


Figure 6.2: *A special SISO multirate sampled-data system with the "vectorized" excitation*

and

$$\int_{-\pi}^{\pi} \frac{\{[\underline{P}(e^{j\omega}) - \underline{\hat{P}}(e^{j\omega})]\phi_{\underline{u}}[\underline{P}^T(e^{j\omega}) - \underline{\hat{P}}^T(e^{j\omega})] + |N_{nT}(e^{j\omega})|^2 \phi_e\}}{\left|\hat{N}_{nT}(e^{j\omega})\right|^2} d\omega = \int_{-\pi}^{\pi} \frac{\{\sigma_n + |N_{nT}(e^{j\omega})|^2 \phi_e\}}{\left|\hat{N}_{nT}(e^{j\omega})\right|^2}$$

where σ_n is the smallest singular value of $\phi_{\underline{u}}$. We know that

$$\underline{u}(k) = \begin{bmatrix} u(nk) \\ \vdots \\ u(nk+n-1) \end{bmatrix};$$

if

$$u(nk) = \cdots = u(nk+n-1)$$

for most of the time ($k = 1, \cdots, M$), then there is not enough excitation, and $\sigma_n \to 0$. When there is not enough excitation, those methods in the polynomial domain may give wrong results. In order to avoid this, we have developed a "vectorized" excitation method: The objective is to design a vector input sequence $\underline{u}(k)$ whose components are independent to each other, and inverse lift $\underline{u}(k)$ by L_n^{-1} before it is inserted into the multirate process. The block diagram of this method is shown in Figure 6.2. The quantity $\phi_{\underline{u}}$ for the "vectorized" excitation is a diagonal matrix, and with a good signal to noise ratio, good identification results are guaranteed.

6.4 Industrial Case Study

In this section we will use an industrial case study to validate the results presented in this chapter.

The octane content is an important quality criterion in the gasoline production units. The continuous catalytic reforming (CCR) unit is responsible for upgrading

the research octane number (RON) of a naphtha feed stream. The increase in RON is accomplished by conversion of a naphthene and paraffin material to aromatic. Therefore, it is an indication of the severity of the CCR reaction. The CCR reactor consists of 4 beds with continuous catalyst circulation and regeneration. Because the reforming reactions are largely endothermic, heat must be added to the reactor feed prior to entering each bed. The heat is provided by a 4 cell balanced draft fired heater with a common convection section. The reactor bed volumes are different and increase with each successive bed. Currently, a weighted average inlet temperature (WAIT) QDMC application is used to control the reactor severity.

There are also many other variables, affecting the RON, inside the unit such as loop pressure, reactor temperatures and outside the unit such as the feed composition changes. For instance, as the amount of feed precursors is decreased, the RON will also decrease for a given severity. If this change is measured, increasing the reactor weighted average inlet temperature (WAIT) can offset the decrease in RON. By implementing RON control, the effect of process disturbances can be minimized. The WAIT can also be manipulated to compensate for other critical operating variables such as decreased catalyst activity.

Usually the octane is sampled and tested in the plant laboratory on a daily basis. Good octane control needs online and frequent measurements of the octane content which requires expensive analytical equipment. There are limited technologies in the market but all require large capital investment and extensive maintenance efforts. As a tradeoff of performance and investment, an octane GC analyzer was installed in the process environment to measure the composition of the CCR product stream and the octane is validated and calculated for online measurement. It provides the octane reading every 2.5 hours. Though it seems slow for a typical control application, it is certainly a big step forward towards plant optimization. Currently, the WAIT target is set by operators based on the slow sampled octane measurement and the desired octane requirement, i.e. the plant runs under 'open loop'.

The main objective of this control application is to close the loop for octane control. Due to the slow sampling rate of the analyzer, an inferential octane model has to be developed to allow the control application to make moves at a faster rate. Reactor WAIT is the manipulated variable and is desired to be adjusted every 30 minutes. All other disturbance variable measurements are also available to estimate the octane at inter-sampling intervals. Therefore, the first step is to identify the dynamic models from all input variables to the output octane variable; clearly this is a multirate model identification problem.

For the purpose of convenience, we use y to represent the output, u to represent

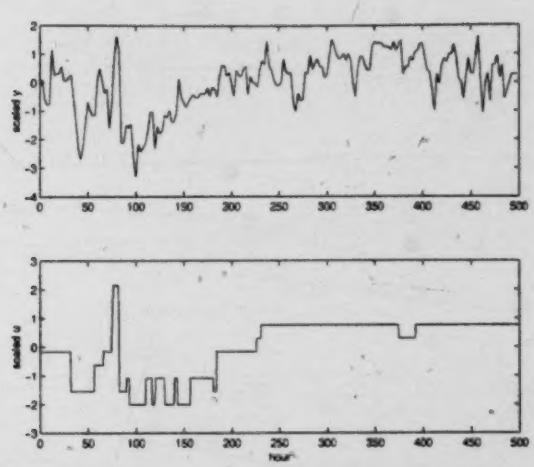


Figure 6.3: *Output and manipulated variable measurements*

the manipulated variable, and d_i $(i = 1, \cdots 7)$ to represent the disturbances. A total of 205 output data are collected at a rate of 150 minutes. The corresponding input/disturbances are collected at a fast rate of 30 minutes. The detrended output/manipulated variables are shown in Figure 6.3. Three disturbances are shown in Figure 6.4 (the reason why only three disturbances are shown will be stated later).

There are 1 output and 8 inputs (manipulated variable and disturbances), and the ratio between the output sampling interval and control interval is 5; so if we lift the inputs in order to get a single-rate lifted system, there will be 40 inputs in the lifted system. Our objective is to estimate the model of the lifted system and then extract a fast sampled model from the lifted model; clearly it is difficult to estimate the model of the system with 40 inputs from only 200 data points. So before we estimate the model, we ask a question: Do all the inputs affect the output significantly? If some of the inputs do not affect the output much, then we can ignore these inputs and estimate a model between the output and the important inputs. In order to answer the question, we formulate two data sets, the output set Y and the lifted input set U, in the following way:

$$\begin{aligned} Y &= y(5k) \\ U &= \left[\, u(5k) \;\; u(5k+1) \;\; \cdots \;\; u(5k+4) \;\; \cdots \;\; d_7(5k) \;\; \cdots \;\; d_7(5k+4) \,\right]; \end{aligned}$$

knowing canonical correlation analysis (CCA) can quantify the relations between two sets of variables [35], we compute the CCA relations between Y and U. The CCA relations show that only the manipulated variable and 3 other disturbances affect the

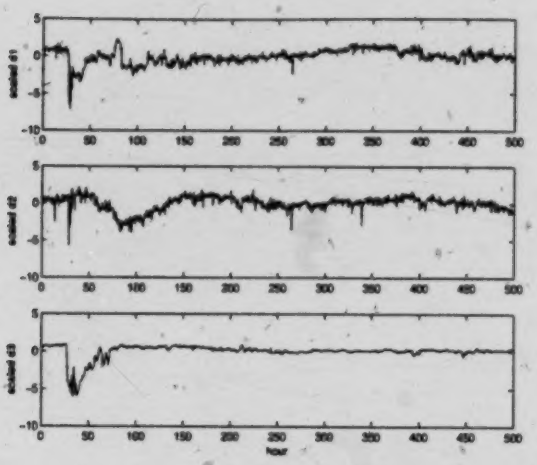


Figure 6.4: *Three disturbances*

output significantly (that is why only three disturbances were shown in Figure 6.4). Hence we estimate the model between the output and 4 inputs (the manipulated variable and 3 disturbances which affect the output significantly) giving us 20 inputs to the lifted system.

We assume that the system can be represented by an ARX model [55], and then formulate a quadratic optimization problem which satisfies the causality constraint. We solve the optimization problem and get the lifted model, and then extract a fast sampled model from this lifted model. Because the data set is too small, again we use the same data set to validate the estimated fast model. Ideally another data set should be used to validate the model if there are enough data. The comparison between the output measurement and the prediction of the model is shown in Figure 6.5.

The modeling result is reasonable with variance of the prediction error being 0.04. The models can be further improved by collecting more data with more input excitations. As mentioned above, a data set of 205 data points is relatively small for estimating a model with 20 inputs. Figures 6.3 and 6.4 also show that the changes in both the manipulated variable and disturbance variables are not rich enough. There are other unmeasured disturbance/noise in the data set as well. So, all these factors would affect the quality of the model obtained.

The inferential models were implemented into a QDMC control application with a control interval of 30 minutes. The controller is multi-rate which makes optimal moves at a fast rate of 30 minutes while the output variable is measured every 2.5

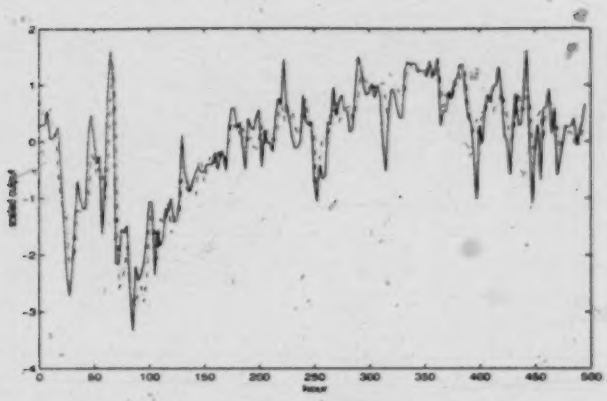


Figure 6.5: *The output measurement (solid line) and the prediction of the fast model (dashdot line*

hours. When the slow sampled output measurement is available, it is used for bias or feedback correction. If not, the model gives an output estimate for control. After one week of online operation, the control variance is shown in Figure 6.6 and compared with manual (open-loop) control in the past. It can be seen that the controller has reduced the output variance by 40%. Therefore, significant economic benefits have been achieved.

6.5 Conclusion

In this chapter we have shown by a theoretical analysis and industrial application, that a fast single-rate transfer function model in which the sampling interval is the same as the control interval of the multirate system can be estimated from multirate data collected from a special but practical class of multirate systems where the output sampling period is an integer multiple of the control period. The multirate esitmator was used under QDMC control of the research octane number in an industrial application with significant control improvement. Compared with multirate identification in the state-space domain, multirate identification in the polynomial domain has two advantages: there is no time-delay-related observability problem, and it is easier to compute the fast-sampled model from the lifted model in the polynomial domain.

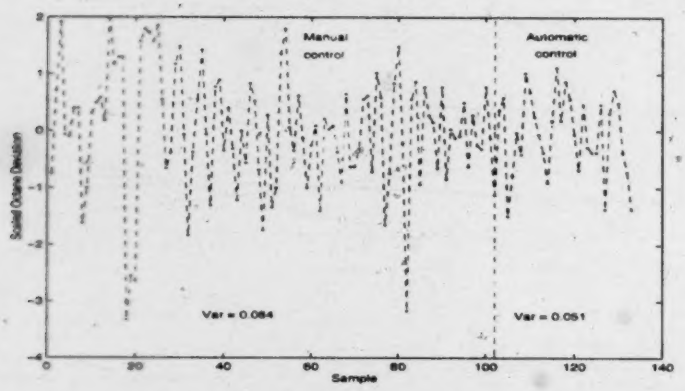


Figure 6.6: *Comparison of performance before and after implementing automatic control*

Chapter 7

System Identification of Closed-loop Multirate Systems

7.1 Introduction

In multirate systems, some of the output sampling intervals may be so large that some loops can not be closed. The CCR reactor we studied in Chapter 6 is an example of such a system. In the industrial case study in Chapter 6, the CCR unit is only an open-loop system. But other situations do exist where some of the loops are closed for economic or safety reasons. Most of the traditional subspace-based schemes can not be applied to such systems which have both open-loop and closed-loop processes. The MOESP in the EIV case developed by Chou and Verhaegen can be applied to both open and closed-loop systems assuming that there is a first-order delay in the feedback loop [13]. Multirate systems with both open loops and closed loops are not uncommon in the chemical industry. Clearly the multirate inferential control scheme can close the loops for such multirate systems, and hence improve the performance of such systems [57, 63, 47]. It is well known that the key to the multirate inferential control strategy is the fast-sampled process model, so it is of great importance to develop subspace-based schemes for identifying fast-sampled models of such multirate systems.

Identification with output feedback has received much attention in the transfer function domain. The closed-loop identification methods in the transfer function domain fall into two main groups: the direct approach and the indirect approach. The direct approach applies the basic prediction error method in a straightforward manner: use the output of the process and the input to the plant in the same way as for open loop identification. Assuming that the regulator is known, the indirect approach identifies the closed-loop system from reference inputs or dither signals added

to the control signal to the process outputs, and retrieves the process model from the closed-loop model. The direct identification methods work well regardless of the complexity of the regulator, and do not require special algorithms; but the approach needs good noise models which are sometimes very difficult to parameterize, hence not all the open-loop methods can be applied to the direct approach. All the open-loop identification methods can be applied to the indirect approach, because the indirect approach basically handles an open-loop problem: the major disadvantage with the indirect approach is that it requires the information of the regulator. Gustavsson and co-workers [31] have surveyed some problems in closed-loop identification (identifiability and accuracy etc.) in the polynomial domain. Forssell and Ljung [22] have studied the statistical properties of the closed-loop identification in the prediction error framework. Extensive work in the area of closed-loop identification in the polynomial domain has been summarized in Ljung's book [55].

In the tutorial introduction of the N4SID and MOESP in Chapter 2, it is evident that one of the basic assumptions of the two algorithms is that the inputs are independent of the disturbances. The subspace methods typically will not give consistent estimates when applied to closed-loop data [55]. This is one of the major disadvantages of the subspace methods. Researchers have tried to apply subspace methods to closed-loop identification in different ways; for example, Verhaegen [78] has applied the MOESP algorithm to closed-loop identification: in his work, he used the indirect approach. Subspace-based closed-loop identification is still a relatively new and active area.

Most of the subspace methods have been developed for the processes which can be represented by ARMAX models (auto-regressive moving-average and exogenous). But ARMAX models do not adequately represent general class of processes [55]. Subspace-based methods which can be applied to the representation of a general class of processes, therefore they are more practical than the ARMAX model.

In this chapter we will present a subspace-based identification method which consider the most general class of processes, and give consistent estimates of the deterministic parts of processes under both open-loop and closed-loop conditions. When applied to a multirate system where some loops are closed, this method can estimate the lifted model for the multirate system; given some conditions, this method can even compute the fast sampled model directly from multirate data.

This chapter is organized as follows. Section 2 gives state-space models for an ARMAX process and a general process, and discusses the difference of the two models. Methods to estimate the state-space model of the deterministic parts of the general class of single-rate systems under open-loop and closed-loop conditions are outlined

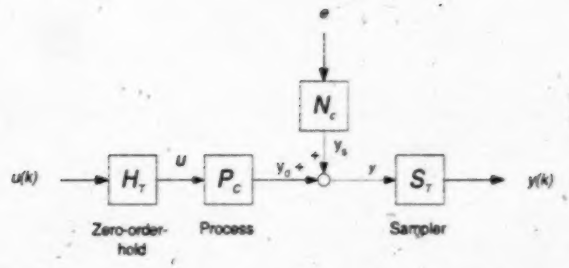


Figure 7.1: *Block diagram of a general combined deterministic-stochastic sampled-data system*

in sections 3 and 4, respectively. In section 5, the single-rate method in sections 3 and 4 will be extended to identify multirate systems, where some loops are closed. The effect of the noise/disturbance on the estimation results is discussed in section 6. The consistency aspects of the proposed algorithm are discussed in section 7 followed by both simulation and experimental examples to illustrate application of the proposed algorithm in section 8. Concluding remarks are given in section 9.

7.2 Preliminaries

In this section we describe the linear time-invariant (LTI) system that is the subject of this study. The block diagram of the open-loop system we will consider in this chapter is shown in Figure 7.1.

The process in Figure 7.1 is a combined deterministic-stochastic LTI sampled-data system. Here, P_c is a continuous-time process with an additive noise; the noise is generated by a continuous-time model N_c with a white noise input e; the output of P_c is corrupted by that of N_c, and is sampled by a sampler S_T with a period T, yielding the sampled output $y(k)$; the input to P_c, u, is generated by a zero-order hold with a period T processing the input sequence $u(k)$. Define

$$y_d = P_c u, \qquad y_s = N_c e,$$

so

$$y = y_d + y_s = P_c u + N_c e.$$

Let P represent the discrete deterministic system, $S_T P_c H_T$, and N the discrete stochastic system, $S_T N_c H_T$. Then

$$y(k) = Pu(k) + Ne(k).$$

Most of the subspace-based identification algorithms, such as CVA [45], N4SID VanOverschee:7 and MOESP [79], assume that the sampled combined system can be represented by such a model as follows,

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Ke(k) \\ y(k) &= Cx(k) + Du(k) + e(k), \end{aligned}$$

where K is the so called Kalman gain matrix. This model is equivalent to a transfer function model:

$$y(k) = [C(zI - A)^{-1}B + D]u(k) + [C(zI - A)^{-1}K + I]e(k). \tag{7.1}$$

Clearly equation (7.1) can be written as

$$y(k) = \frac{E(z)}{F(z)}u(k) + \frac{G(z)}{F(z)}e(k),$$

which is exactly an ARMAX model [55]. The ARMAX model assumes that the deterministic and stochastic systems have the same poles. But in practice, this assumption may not always hold. If the deterministic and stochastic systems have different poles, then the combined sampled-data system can only be represented by

$$y(k) = \frac{E(z)}{F(z)}u(k) + \frac{G(z)}{R(z)}e(k). \tag{7.2}$$

Equation (7.2) is a Box-Jenkins model (BJ) [6] which is more general than an ARMAX model. In this chapter we focus on processes represented by BJ models, so we assume that the system in Figure 7.1 is a BJ process. We will develop the discrete state-space model for the BJ process in Figure 7.1 via the following steps: Assuming that the state-space model of P is

$$\begin{aligned} x_d(k+1) &= A_d x_d(k) + B_d u(k) \\ y_d(k) &= C_d x_d(k) + D_d u(k), \end{aligned}$$

and the state-space model of the discretized stochastic process N is

$$\begin{aligned} x_s(k+1) &= A_s x_s(k) + B_s e(k) \\ y_s(k) &= C_s x_s(k) + D_s e(k), \end{aligned}$$

where $e(k)$ is the discretized white noise sequence. The combined sampled-data system can then be expressed by the following discrete state-space model,

$$\begin{aligned} x(k+1) &= Ax(k) + B_1 u(k) + B_2 e(k) \\ y(k) &= Cx(k) + D_d u(k) + D_s e(k), \end{aligned} \tag{7.3}$$

with

$$x = \begin{bmatrix} x_d \\ x_s \end{bmatrix}, \; y = y_d + y_s, \; A = \begin{bmatrix} A_d & 0 \\ 0 & A_s \end{bmatrix},$$
$$B_1 = \begin{bmatrix} B_d \\ 0 \end{bmatrix}, B_2 = \begin{bmatrix} 0 \\ B_s \end{bmatrix}, C = \begin{bmatrix} C_d & C_s \end{bmatrix},$$

where $x_d \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $u \in \mathbb{R}^l$, $e \in \mathbb{R}^s$, $\mathbb{R}^n$ denoting the space of $n \times 1$ vectors, etc.

7.3 Identification of Open-loop Single-rate Systems

The objective of this section is to develop a method for estimating the state-space model of the deterministic system, $S_T P_c H_T$, in Figure 7.1, from the input-output measurements under open-loop condition. Assume that the pair (C_d, A_d) is observable, the combined system is stable, the input $u(k)$ and the output $y(k)$ are measured. Define

$$y_\alpha(k) \triangleq \begin{bmatrix} y(k) \\ y(k+1) \\ \vdots \\ y(k+\alpha-1) \end{bmatrix},$$

where α is a positive integer larger than n which is the dimension of the deterministic process. Similarly we can define $u_\alpha(k)$ and $e_\alpha(k)$. After some straightforward algebraic manipulations and recursive substitutions, we can get the following equation from equation (7.3),

$$y_\alpha(k) = \Gamma_\alpha^d x_d(k) + \Gamma_\alpha^s x_s(k) + H_\alpha^d u_\alpha(k) + H_\alpha^s e_\alpha(k), \tag{7.4}$$

where

$$\Gamma_\alpha^d = \begin{bmatrix} C_d \\ C_d A_d \\ \vdots \\ C_d (A_d)^{\alpha-1} \end{bmatrix}$$

and

$$H_\alpha^d = \begin{bmatrix} D_d & 0 & \cdots & 0 \\ C_d B_d & D_d & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C_d (A_d)^{\alpha-2} B_d & C_d (A_d)^{\alpha-3} B_d & \cdots & D_d \end{bmatrix}.$$

Γ_α^s can be obtained by replacing A_d and C_d in Γ_α^d by A_s and C_s, respectively. Similarly H_α^s can be obtained by replacing A_d, B_d, C_d and D_d in H_α^d by A_s, B_s, C_s and D_s, respectively. Γ_α^d is the so called extended observability matrix of the deterministic

system, and H_α^d is the impulse response matrix. Define the output and state Hankel matrices as follows,

$$Y_{\alpha,J}(k) \triangleq [\ y_\alpha(k)\ \ y_\alpha(k+1)\ \ \cdots\ \ y_\alpha(k+J-1)\ ] \tag{7.5}$$

$$X_{1,J}^d(k) \triangleq [\ x_d(k)\ \ x_d(k+1)\ \ \cdots\ \ x_d(k+J-1)\ ], \tag{7.6}$$

where J is the observation number. Similarly we can define $U_{\alpha,J}(k)$, $E_{\alpha,J}(k)$ and $X_{1,J}^s(k)$. Hence we can write an equation similar to equation (7.4):

$$Y_{\alpha,J}(k+\beta) = \Gamma_\alpha^d X_{1,J}^d(k+\beta) + \Gamma_\alpha^s X_{1,J}^s(k+\beta) + H_\alpha^d U_{\alpha,J}(k+\beta) + H_\alpha^s E_{\alpha,J}(k+\beta), \tag{7.7}$$

where β is a positive integer larger than or equal to α. If we can find an instrumental variable which is independent of the noise $e(k)$, but correlated with the state $x(k)$ and input $u(k)$, then we can remove the effect of the noise [68, 12]. Usually we can assume that the excitation input $u(k)$ is independent of the noise $e(k)$, so $u(k)$ can be a natural choice as the instrumental variable. Post-multiplying both sides of equation (7.7) by $\frac{1}{J}U_{\alpha,J}^T(k)$ gives:

$$\begin{aligned}\frac{1}{J}Y_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k) &= \frac{1}{J}\{\Gamma_\alpha^d X_{1,J}^d(k+\beta)U_{\alpha,J}^T(k) + \Gamma_\alpha^s X_{1,J}^s(k+\beta)U_{\alpha,J}^T(k) \\ &\quad + H_\alpha^d U_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k) + H_\alpha^s E_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k)\}\end{aligned} \tag{7.8}$$

It follows from the definitions of $E_{\alpha,J}(k+\beta)$ and $U_{\alpha,J}(k)$ that

$$\lim_{J\to\infty}\frac{1}{J}E_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k) = \begin{bmatrix} \gamma_{eu}(\beta) & \cdots & \gamma_{eu}(\beta-\alpha+1) \\ \vdots & \ddots & \vdots \\ \gamma_{eu}(\beta-\alpha+1) & \cdots & \gamma_{eu}(\beta) \end{bmatrix},$$

where

$$\gamma_{eu}(\beta) = \lim_{J\to\infty}\frac{1}{J}\sum_{k=1}^{J} e(k+\beta)u(k).$$

The state-space model of the stochastic part of the process is

$$\begin{aligned} x_s(k+1) &= A_s x_s(k) + B_s e(k) \\ y_s(k) &= C_s x_s(k) + D_s e(k). \end{aligned}$$

The equivalent transfer function model is

$$x_s(k) = (zI - A_s)^{-1}B_s e(k), \tag{7.9}$$

where z is the forward shift operator. Equation (7.9) and the definitions of $X_{1,J}^s(k+\beta)$ and $U_{\alpha,J}(k)$ give

$$\lim_{J\to\infty}\frac{1}{J}X_{1,J}^s(k+\beta)U_{\alpha,J}^T(k) = (zI - A_s)^{-1}B_s\ [\ \gamma_{eu}(\beta)\ \ \cdots\ \ \gamma_{eu}(\beta-\alpha+1)\ ].$$

Since e is independent of u, we have

$$\gamma_{eu}(k) = 0, \ k = 0, 1, 2, ...$$

If the stochastic system is stable, then Γ_α^s and H_α^s are finite. Hence

$$\lim_{J\to\infty} \frac{1}{J}\Gamma_\alpha^s X_{1,J}^s(k+\beta)U_{\alpha,J}^T(k) = 0,$$

and

$$\lim_{J\to\infty} \frac{1}{J}H_\alpha^s E_{\alpha,J}^s(k+\beta)U_{\alpha,J}^T(k) = 0.$$

Clearly when $J \to \infty$, equation (7.8) can be reduced to:

$$\lim_{J\to\infty} \frac{1}{J}Y_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k) = \frac{1}{J}[\Gamma_\alpha^d X_{1,J}^d(k+\beta) + H_\alpha^d U_{\alpha,J}(k+\beta)]U_{\alpha,J}^T(k). \quad (7.10)$$

If J is not infinitely large, then:

$$\frac{1}{J}Y_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k) = \frac{1}{J}[\Gamma_\alpha^d X_{1,J}^d(k+\beta) + H_\alpha^d U_{\alpha,J}(k+\beta)]U_{\alpha,J}^T(k) + \Xi(J),$$

where

$$\Xi(J) = \frac{1}{J}[H_\alpha^s E_{\alpha,J}(k+\beta) + \Gamma_\alpha^s X_{1,J}^s(k+\beta)]U_{\alpha,J}^T(k).$$

In sections 3, 4 and 5, we assume that J is infinitely large. The case when J is finite will be discussed in section 6. In the remaining part of this section, we will introduce a method to estimate the state-space model, (A_d, B_d, C_d, D_d), based on equation (7.10).

7.3.1 Estimation of A_d and C_d

In practice, input and output can be measured, so $Y_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k)$ and $U_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k)$ in equation (7.10) are available. But (7.10) is not a linear equation of the system matrices, (A_d, B_d, C_d, D_d). One possible way to compute (A_d, B_d, C_d, D_d) is to calculate Γ_α^d first. This can be done by getting rid of the last part, $H_\alpha^d U_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k)$, in equation (7.10). Clearly a matrix which is correlated to $Y_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k)$ but orthogonal to $U_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k)$ can do the job. In order to generate such a matrix, we first formulate a matrix

$$\frac{1}{J}\begin{bmatrix} U_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k) \\ Y_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k) \end{bmatrix},$$

and then perform the following QR factorization [46]

$$\frac{1}{J}\begin{bmatrix} U_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k) \\ Y_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k) \end{bmatrix} = \begin{bmatrix} R_{11}^J & 0 \\ R_{21}^J & R_{22}^J \end{bmatrix}\begin{bmatrix} (Q_1^J)^T \\ (Q_2^J)^T \end{bmatrix}.$$

Being correlated with $Y_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k)$ and independent of $U_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k)$, Q_2^J is the matrix we need. Post-multiplying both sides of equation (7.10) by Q_2^J gives

$$[R_{21}^J(Q_1^J)^T + R_{22}^J(Q_2^J)^T]Q_2^J = \frac{1}{J}\Gamma_\alpha^d X_{1,J}^d(k+\beta)U_{\alpha,J}^T(k)Q_2^J + H_\alpha^d R_{11}^J(Q_1^J)^T Q_2^J. \quad (7.11)$$

Because $(Q_1^J)^T Q_2^J = 0$ and $(Q_2^J)^T Q_2^J = I$, equation (7.11) becomes

$$\lim_{J\to\infty} R_{22}^J = \frac{1}{J}\Gamma_\alpha^d X_{1,J}^d(k+\beta)U_{\alpha,J}^T(k)Q_2^J. \quad (7.12)$$

Define

$$\Phi_J \triangleq \frac{1}{J}X_{1,J}^d(k+\beta)U_{\alpha,J}^T(k)Q_2^J.$$

The singular value decomposition (SVD) [46] of R_{22}^J yields

$$R_{22}^J = \begin{bmatrix} U_1^J & U_2^J \end{bmatrix} \begin{bmatrix} S_1^J & 0 \\ 0 & S_2^J \end{bmatrix} \begin{bmatrix} (V_1^J)^T \\ (V_2^J)^T \end{bmatrix},$$

where S_1^J contains the n largest singular values, U_1^J contains the first n columns of the left singular vector matrix, and $(V_1^J)^T$ contains the first n rows of the right singular vector matrix. From equation (7.12) we know that if J goes to infinity, then the rank of matrix R_{22}^J is no larger than the rank of matrix Γ_α^d, and this means that the rank of the matrix $R_{22}^J \leq n$. From the assumption we know that $\alpha > n$, so

$$\begin{aligned} \lim_{J\to\infty} S_2^J &= 0 \\ \lim_{J\to\infty} R_{22}^J &= U_1^J S_1^J (V_1^J)^T. \end{aligned} \quad (7.13)$$

From equations (7.12) and (7.13) it follows that

$$\lim_{J\to\infty} \Gamma_\alpha^d \Phi_J = U_1^J S_1^J (V_1^J)^T, \quad (7.14)$$

where $\Gamma_\alpha^d \in \mathbb{R}^{\alpha m \times n}$, $\Phi_J \in \mathbb{R}^{n \times \alpha m}$, $U_1^J S_1^J \in \mathbb{R}^{\alpha m \times n}$, and $(V_1^J)^T \in \mathbb{R}^{n \times \alpha m}$ ($\mathbb{R}^{\alpha m \times n}$ denotes the space of $\alpha m \times n$ matrices). Then multiplying both sides of equation (7.14) by V_1^J gives

$$\lim_{J\to\infty} U_1^J S_1^J = \Gamma_\alpha^d \Phi_J V_1^J. \quad (7.15)$$

The state-space model of the deterministic process can also be expressed as $(\Lambda A_d \Lambda^{-1}, \Lambda B_d, C_d\Lambda^{-1}, D_d)$, and the extended observability matrix as $\Gamma_\alpha^d \Lambda^{-1}$, with Λ being a full rank square matrix. So if $\Phi_J V_1^J$ is a full rank matrix (we will prove this in section 5), then $U_1^J S_1^J$ is the extended observability matrix. In practice J can not be infinitely large, and the corresponding S_2^J is typically non-zero and contains noise information. But the components in S_1^J will be significantly larger than those in S_2^J when the signal

to noise ratio is large (this will be discussed in section 6). Matrices A_d and C_d can be computed from $U_1^J S_1^J$; one of the simplest methods for estimating (A_d, C_d) is to take the first m rows of $U_1^J S_1^J$ as matrix C_d, then compute the matrix A_d through the following steps. First formulate two matrices,

$$\Gamma_1 = \begin{bmatrix} C_d \\ C_d A_d \\ \vdots \\ C_d A_d^{n-1} \end{bmatrix}, \quad \Gamma_2 = \begin{bmatrix} C_d A_d \\ C_d A_d^2 \\ \vdots \\ C_d A_d^n \end{bmatrix},$$

from the estimated extended observability matrix $U_1^J S_1^J$. Clearly

$$\Gamma_2 = \Gamma_1 A_d,$$

so A_d can be computed as

$$A_d = (\Gamma_1^T \Gamma_1)^{-1} \Gamma_1^T \Gamma_2.$$

If the pair $(C_d,\ A_d)$ are observable, then matrix A_d can be computed uniquely; and this is a least square solution.

7.3.2 Estimation of B_d and D_d

Define

$$(\Gamma_\alpha^d)^\perp \triangleq I - \Gamma_\alpha^d [(\Gamma_\alpha^d)^T \Gamma_\alpha^d]^{-1} (\Gamma_\alpha^d)^T.$$

After pre-multiplying both sides of equation (7.10) by $(\Gamma_\alpha^d)^\perp$, we have

$$\lim_{J\to\infty} \frac{1}{J} (\Gamma_\alpha^d)^\perp Y_{\alpha,J}(k+\beta) U_{\alpha,J}^T(k) = \frac{1}{J} (\Gamma_\alpha^d)^\perp H_\alpha^d U_{\alpha,J}(k+\beta) U_{\alpha,J}^T(k),$$

and

$$\lim_{J\to\infty} \frac{1}{J} (\Gamma_\alpha^d)^\perp Y_{\alpha,J}(k+\beta) U_{\alpha,J}^T(k) [\frac{1}{J} U_{\alpha,J}(k+\beta) U_{\alpha,J}^T(k)]^{-1} = (\Gamma_\alpha^d)^\perp H_\alpha^d. \tag{7.16}$$

Define

$$H_1 \triangleq \lim_{J\to\infty} \frac{1}{J} (\Gamma_\alpha^d)^\perp Y_{\alpha,J}(k+\beta) U_{\alpha,J}^T(k) [\frac{1}{J} U_{\alpha,J}(k+\beta) U_{\alpha,J}^T(k)]^{-1}.$$

Here $\frac{1}{J} U_{\alpha,J}(k+\beta) U_{\alpha,J}^T(k)$ is assumed to be invertable, and this should be guaranteed when the excitation being designed. Matrices Γ_α^d, $Y_{\alpha,J}(k+\beta)$, $U_{\alpha,J}(k)$ and $U_{\alpha,J}(k+\beta)$ are known, so the left hand side of equation (7.16), H_1, can be computed. Observing the structure of H_α^d given earlier, we have

$$\begin{bmatrix} D_d \\ C_d B_d \\ \vdots \\ C_d (A_d)^{\alpha-2} B_d \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & C_d \\ \vdots & \vdots \\ 0 & C_d (A_d)^{\alpha-2} \end{bmatrix} \begin{bmatrix} D_d \\ B_d \end{bmatrix}.$$

Define two matrices as follows:

$$H_R = \begin{bmatrix} (\Gamma_\alpha^d)^\perp \begin{bmatrix} I & 0 \\ 0 & C_d \\ \vdots & \vdots \\ 0 & C_d(A_d)^{\alpha-2} \end{bmatrix} \\ (\Gamma_\alpha^d)^\perp \begin{bmatrix} I & 0 \\ \vdots & \vdots \\ 0 & C_d(A_d)^{\alpha-3} \end{bmatrix} \\ \vdots \end{bmatrix}$$

and

$$H_L = \begin{bmatrix} \text{first block column of } H_1 \\ \text{the last } (\alpha-1)m \text{ rows of the second block column of } H_1 \\ \vdots \\ \text{the last } m \text{ rows of the last block column of } H_1 \end{bmatrix}.$$

Consequently, it follows from equation (7.16) that

$$H_L = H_R \begin{bmatrix} D_d \\ B_d \end{bmatrix},$$

which results in

$$\begin{bmatrix} D_d \\ B_d \end{bmatrix} = (H_R^T H_R)^{-1} H_R^T H_L.$$

A similar method to compute matrices (A_d, B_d, C_d, D_d) was first developed by Verhaegen in the MOESP approach [80, 81].

7.4 Identification of Closed-loop Single-rate Systems

Consider the closed-loop sampled-data system in Figure 7.2. Here $r(k)$ is the setpoint, $f(k)$ is the excitation sequence, G is the controller and $u(k)$ is the output of the controller; $u(t)$, the input into the process P_c, is generated by a zero-order hold with period T processing the discrete sequence $u(k)$; the output of P_c is corrupted by that of N_c, and is sampled by S_T, yielding the sampled output $y(k)$. Our method requires the assumption that $f(k)$ is independent of the disturbance. For simplicity we assume that $r(k)$ is 0. The state-space models of sampled-data systems, P and N, have been given in section 3. The discrete system relating $u(k)$ to $y(k)$ can be expressed by a state-space model,

$$\begin{aligned} x(k+1) &= Ax(k) + B_1 u(k) + B_2 e(k) \\ y(k) &= Cx(k) + D_d u(k) + D_e e(k), \end{aligned}$$

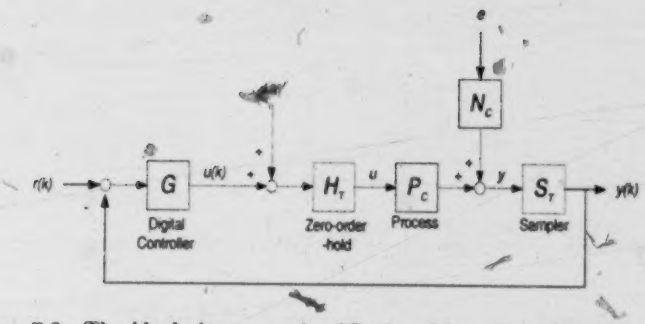


Figure 7.2: *The block diagram of a SR closed-loop sampled-data system*

where $e(k)$ is the discretized white noise sequence. Assume that (C_d, A_d) is an observable pair, and that both the open-loop and the closed-loop systems are stable. The objective is to estimate the state-space model (A_d, B_d, C_d, D_d) of the system from the measured $u(k)$, $f(k)$ and $y(k)$. It is routine to get an equation similar to equation (7.7):

$$Y_{\alpha,J}(k+\beta) = \Gamma_\alpha^d X_{1,J}^d(k+\beta) + \Gamma_\alpha^s X_{1,J}^s(k+\beta) + H_\alpha^d U_{\alpha,J}(k+\beta) + H_\alpha^s E_{\alpha,J}(k+\beta). \quad (7.17)$$

Clearly if we can find an instrumental variable which is independent of the noise sequence $e(k)$, but correlated with the state $x(k)$ and input $u(k)$, then we can remove the effect of the noise. Substituting

$$u(k) = (I+GP)^{-1}GPf(k) - (I+GP)^{-1}GNe(k) \quad (7.18)$$

into (7.17) gives

$$Y_{\alpha,J}(k+\beta)F_{\alpha,J}^T(k) = \Gamma_\alpha^d X_{1,J}^d(k+\beta) + H_\alpha^d(I+GP)^{-1}GPF_{\alpha,J}(k+\beta) + \mathbb{E} \quad (7.19)$$

where

$$\mathbb{E} = \Gamma_\alpha^s X_{1,J}^s(k+\beta) + H_\alpha^s E_{\alpha,J}(k+\beta) - H_\alpha^d(I+GP)^{-1}GNE_{\alpha,J}(k+\beta).$$

Being correlated with $e(k)$, $u(k)$ can not be used as an instrumental variable. Equation (7.19) indicates that $f(k)$ can be the instrumental variable if it is independent of $e(k)$. At the beginning of this section we assumed that $f(k)$ is independent of $e(k)$. Hence we can choose $f(k)$ as the instrumental variable. $F_{\alpha,J}(k)$ can be obtained by replacing $y(k)$ in $Y_{\alpha,J}(k)$ with $f(k)$. Post-multiplying both sides of equation (7.19) by $\frac{1}{J}F_{\alpha,J}^T(k)$ gives

$$\lim_{J\to\infty}\frac{1}{J}Y_{\alpha,J}(k+\beta)F_{\alpha,J}^T(k) = \lim_{J\to\infty}\frac{1}{J}[\Gamma_\alpha^d X_{1,J}^d(k+\beta) + H_\alpha^d(I+GP)^{-1}GPF_{\alpha,J}(k+\beta) + \mathbb{E}]F_{\alpha,J}^T(k). \quad (7.20)$$

It follows from the definitions of $E_{\alpha,J}(k+\beta)$ and $F_{\alpha,J}(k)$ that

$$\lim_{J\to\infty}\frac{1}{J}E_{\alpha,J}(k+\beta)F_{\alpha,J}^T(k)=\begin{bmatrix}\gamma_{ef}(\beta) & \cdots & \gamma_{ef}(\beta-\alpha+1)\\ \vdots & \ddots & \vdots\\ \gamma_{ef}(\beta-\alpha+1) & \cdots & \gamma_{ef}(\beta)\end{bmatrix},$$

where

$$\gamma_{ef}(\beta)=\lim_{J\to\infty}\frac{1}{J}\sum_{k=1}^{J}e(k+\beta)f(k).$$

Substituting equation (7.9) into $X_{1,J}^s(k+\beta)F_{\alpha,J}^T(k)$ yields

$$\lim_{J\to\infty}\frac{1}{J}X_{1,J}^s(k+\beta)F_{\alpha,J}^T(k)=(zI-A_s)^{-1}B_s\begin{bmatrix}\gamma_{ef}(\beta) & \cdots & \gamma_{ef}(\beta-\alpha-1).\end{bmatrix}$$

Since e is independent of f,

$$\gamma_{ef}(k)=0,\ k=0,1,2,....$$

If the open-loop and closed-loop systems are stable, then Γ_α^s and H_α^s are finite. Hence

$$\lim_{J\to\infty}\frac{1}{J}\Gamma_\alpha^s X_{1,J}^s(k+\beta)F_{\alpha,J}^T(k)=0,\ \lim_{J\to\infty}\frac{1}{J}H_\alpha^s E_{\alpha,J}^s(k+\beta)U_{\alpha,J}^T(k)=0.$$

and

$$\lim_{J\to\infty}\frac{1}{J}H_\alpha^d(I+GP)^{-1}GNE_{\alpha,J}(k+\beta)F_{\alpha,J}^T(k)=0.$$

We can conclude that when J goes to infinity, equation (7.20) reduces to

$$\lim_{J\to\infty}\frac{1}{J}Y_{\alpha,J}(k+\beta)F_{\alpha,J}^T(k)=\lim_{J\to\infty}\frac{1}{J}[\Gamma_\alpha^d X_{1,J}^d(k+\beta)+H_\alpha^d(I+GP)^{-1}F_{\alpha,J}(k+\beta)]F_{\alpha,J}^T(k). \tag{7.21}$$

In order to get rid of the second part on the right hand side of (7.21), we need a matrix which is correlated with $Y_{\alpha,J}(k+\beta)F_{\alpha,J}^T(k)$ but orthogonal to $F_{\alpha,J}(k+\beta)F_{\alpha,J}^T(k)$. Once again we perform the following QR factorization:

$$\frac{1}{J}\begin{bmatrix}F_{\alpha,J}(k+\beta)F_{\alpha,J}^T(k)\\ Y_{\alpha,J}(k+\beta)F_{\alpha,J}^T(k)\end{bmatrix}=\begin{bmatrix}R_{11,c}^J & 0\\ R_{21,c}^J & R_{22,c}^J\end{bmatrix}\begin{bmatrix}(Q_{1,c}^J)^T\\ (Q_{2,c}^J)^T\end{bmatrix}.$$

Because $(Q_{1,c}^J)^TQ_{2,c}^J=0$ and $(Q_{2,c}^J)^TQ_{2,c}^J=I$, post-multiplying both sides of equation (7.21) by $Q_{2,c}^J$ gives

$$\lim_{J\to\infty}R_{22,c}^J=\frac{1}{J}\Gamma_\alpha^d X_{1,J}^d(k+\beta)F_{\alpha,J}^T(k)Q_{2,c}^J. \tag{7.22}$$

Singular value decomposition of $R_{22,c}^J$ yields

$$\begin{aligned}R_{22,c}^J &= U_c^JS_c^J(V_c^J)^T\\ &= \begin{bmatrix}U_{1,c}^J & U_{2,c}^J\end{bmatrix}\begin{bmatrix}S_{1,c}^J & 0\\ 0 & S_{2,c}^J\end{bmatrix}\begin{bmatrix}(V_{1,c}^J)^T\\ (V_{2,c}^J)^T\end{bmatrix},\end{aligned}$$

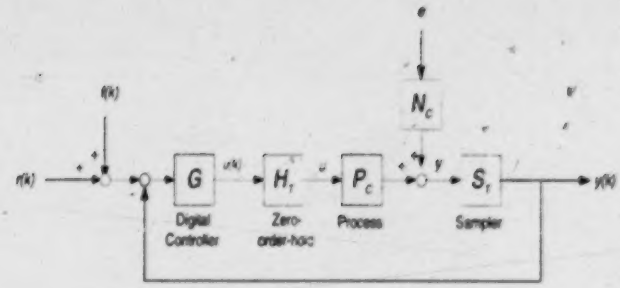


Figure 7.3: *The block diagram of a closed-loop sampled-data system*

where $S_{1,c}^J$ contains the n largest singular values, $U_{1,c}^J$ contains the first n columns of the left singular vector matrix U_c^J, and $(V_{1,c}^J)^T$ contains the first n rows of the right singular vector matrix $(V_c^J)^T$. From equation (7.22) we know that if data length J goes to infinity, the rank of matrix $R_{22,c}^J \leq n$. From the assumption that $\alpha > n$, it follows that

$$\lim_{J\to\infty} U_{1,c}^J S_{1,c}^J = \Gamma_\alpha^d \Lambda_c^d, \tag{7.23}$$

where

$$\Lambda_c^d = \lim_{J\to\infty} \frac{1}{J} X_{1,J}^d(k+\beta) F_{a,J}^T(k) Q_{2,c}^J V_c^J. \tag{7.24}$$

$U_{1,c}^J S_{1,c}^J$ is the extended observability matrix if and only if Λ_c^d is a full rank matrix. In section 5 we will prove that Λ_c^d is a full rank matrix provided that certain conditions are satisfied. We can first calculate matrices A_d and C_d from $U_{1,c}^J S_{1,c}^J$, then it is straightforward to calculate matrices B_d and D_d.

In some cases it is more convenient to add excitation sequence on the setpoint (see Figure 7.3). In such case the control signal $u(k)$ can be expressed as

$$u(k) = (I+GP)^{-1}Gf(k) - (I+GP)^{-1}GNe(k).$$

Since $f(k)$ can be assumed to be independent of $e(k)$, and $f(k)$ is obviously correlated with the state $x(k)$ and input $u(k)$, we can also choose $f(k)$ to be the instrumental variable. The way to estimate the state-space model (A_d, B_d, C_d, D_d) is the same as that discussed in the earlier part of this section.

7.5 Identification of Closed-loop Multirate Systems

In this section we will discuss identification of a special case of multirate systems where all the control intervals are uniform, all the output sampling intervals are

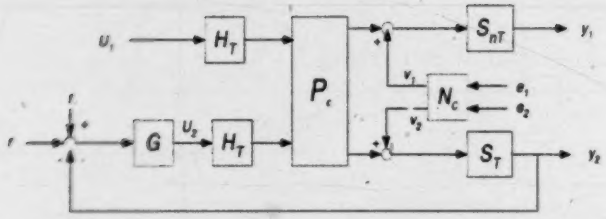


Figure 7.4: *The block diagram of a closed-loop MR sampled-data system*

distinct, but they are all integer multiples of the control interval. In such multirate system some loops are closed. The objective of this section is to develop a method for estimating the fast sampled model where the output is sampled as fast as the input from the measured input-output data. For simplicity, we will use a simple 2-input and 2-output multirate system to illustrate our method. Consider the sampled-data system in Figure 7.4; the continuous process, P_c, is an LTI system with two inputs and two outputs; all the holds in the system have uniform interval, T; one of the output is sampled every period nT, but the other output is sampled every period T. The output with sampling interval nT is manually controlled by u_1; at the same time, the output with sampling interval T is controlled by a controller G.

If we can estimate the discrete model of P_c sampled every interval T, then we can design an inferential controller to control the two outputs. In order to identify the fast model, we have to make a few assumptions: (1) u_1 is independent of e_1 and e_2; (2) an excitation sequence, f, is added on the setpoint of the controller, G, and f is independent of both e_1 and e_2. The multirate system shown in Figure 7.4 is a periodically time-varying system, so it can be transformed into a time-invariant system through the lifting operator. In order to arrive at a single-rate system, we have to lift u_1, u_2 and y_2 by L_n. Define,

$$\underline{y}(k) = \begin{bmatrix} y_1(k) \\ y_2(kn) \\ y_2(kn+1) \\ \vdots \\ y_2(kn+n-1) \end{bmatrix}, \quad \underline{u}(k) = \begin{bmatrix} u_1(kn) \\ u_2(kn) \\ \vdots \\ u_1(kn+n-1) \\ u_2(kn+n-1) \end{bmatrix}, \quad \underline{e}(k) = \begin{bmatrix} e_1(kn) \\ e_2(kn) \\ \vdots \\ e_1(kn+n-1) \\ e_2(kn+n-1) \end{bmatrix}.$$

Let $\underline{P}$ represent the lifted deterministic system, and $\underline{N}$ represent the lifted stochastic system. Hence

$$\underline{y} = \underline{P}\underline{u} + \underline{N}\underline{e}.$$

From the appropriate partitions, we have

$$C_d = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}, \; D_d = \begin{bmatrix} D_1 & D_2 \end{bmatrix},$$

where C_1 relates to y_1, C_2 to y_2, D_1 relates u_1, and D_2 to u_2. The state-space model $\left[\begin{array}{c|c} \underline{A_d} & \underline{B_d} \\ \hline \underline{C_d} & \underline{D_d} \end{array}\right]$ of $\underline{P}$ can be derived by using proposition 1 in Chapter 5:

$$\left[\begin{array}{c|c} \underline{A_d} & \underline{B_d} \\ \hline \underline{C_d} & \underline{D_d} \end{array}\right] = \left[\begin{array}{c|cccc} A_d^n & A_d^{n-1}B_d & A_d^{n-2}B_d & \cdots & B_d \\ \hline C_d & D_d & 0 & \cdots & 0 \\ C_2A_d & C_2B_d & D_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_2A_d^{n-1} & C_2A_d^{n-2}B_d & C_2A_d^{n-3}B_d & \cdots & D_2 \end{array}\right].$$

The state-space model $\left[\begin{array}{c|c} \underline{A_s} & \underline{B_s} \\ \hline \underline{C_s} & \underline{D_s} \end{array}\right]$ of $\underline{N}$ can be similarly derived. It is not difficult to formulate the following equation,

$$\underline{Y}_{\alpha,J}(k+\beta) = \underline{\Gamma}_\alpha^d \underline{X}_{1,J}^d(k+\beta) + \underline{\Gamma}_\alpha^s \underline{X}_{1,J}^s(k+\beta) + \underline{H}_\alpha^d \underline{U}_{\alpha,J}(k+\beta) + \underline{H}_\alpha^s \underline{E}_{\alpha,J}(k+\beta), \quad (7.25)$$

where $\underline{Y}_{\alpha,J}$ can be obtained by replacing y in $Y_{\alpha,J}$ defined in equation (7.5) by $\underline{y}$; $\underline{\Gamma}_\alpha^d$ can be obtained by replacing A_d and C_d in Γ_α^d by $\underline{A_d}$ and $\underline{C_d}$, respectively. Similarly $\underline{X}_{1,J}^d$, $\underline{\Gamma}_\alpha^s$, $\underline{X}_{1,J}^s$, $\underline{H}_\alpha^d$, $\underline{H}_\alpha^s$ and $\underline{E}_{\alpha,J}$ can be obtained. Define

$$\underline{f} = \begin{bmatrix} u_1(0) \\ f(0) \\ \vdots \\ u_1(n-1) \\ f(n-1) \end{bmatrix}, \; \underline{f}_\alpha(k) \triangleq \begin{bmatrix} \underline{f}(k) \\ \vdots \\ \underline{f}(k+\alpha-1) \end{bmatrix},$$

and

$$\underline{F}_{\alpha,J}(k) \triangleq \begin{bmatrix} \underline{f}_\alpha(k) & \underline{f}_\alpha(k+1) & \cdots & \underline{f}_\alpha(k+J-1) \end{bmatrix}.$$

Multiplying both sides of equation (7.25) by $\frac{1}{J}\underline{F}_{\alpha,J}^T(k)$ yields

$$\lim_{J\to\infty} \frac{1}{J}\underline{Y}_{\alpha,J}(k+\beta)\underline{F}_{\alpha,J}^T(k) = \lim_{J\to\infty} \frac{1}{J}[\underline{\Gamma}_\alpha^d \underline{X}_{1,J}^d(k+\beta) + \underline{H}_\alpha^d \underline{U}_{\alpha,J}(k+\beta)]\underline{F}_{\alpha,J}^T(k). \quad (7.26)$$

It is straightforward to calculate $\underline{\Gamma}_\alpha^d$ from equation (7.26), and then compute ($\underline{A_d}$, $\underline{B_d}$, $\underline{C_d}$, $\underline{D_d}$); then with the techniques discussed in Chapter 5, we can extract matrices

$(A_d,\ B_d,\ C_d,\ D_d)$ from matrices $(\underline{A_d},\ \underline{B_d},\ \underline{C_d},\ \underline{D_d})$. But the question that naturelly arises is: can we calculate matrices, $A_d,\ B_d,\ C_d,\ D_d$, directly from equation (7.26). The answer is in the affirmative under certain conditions: Notice that

$$\underline{\Gamma_\alpha^d} = \begin{bmatrix} \underline{C_d} \\ \underline{C_d A_d} \\ \vdots \\ \underline{C_d (A_d)^{\alpha-1}} \end{bmatrix} = \begin{bmatrix} C_d \\ C_2 A_d \\ \vdots \\ C_2 A_d^{n-1} \\ C_d A_d^n \\ C_2 A_d^{n+1} \\ \vdots \\ C_2 A_d^{n\alpha-1} \end{bmatrix}.$$

At first we take the first α rows of $\underline{\Gamma_\alpha^d}$ as matrix C_d, and formulate the following two matrices,

$$\Gamma_R = \begin{bmatrix} C_2 \\ C_2 A_d \\ \vdots \\ C_2 A_d^{n-1} \\ C_2 A_d^n \\ C_2 A_d^{n+1} \\ \vdots \\ C_2 A_d^{n\alpha-2} \end{bmatrix},\ \Gamma_L = \begin{bmatrix} C_2 A_d \\ C_2 A_d^2 \\ \vdots \\ C_2 A_d^{n-1} \\ C_2 A_d^n \\ C_2 A_d^{n+1} \\ \vdots \\ C_2 A_d^{n\alpha-1} \end{bmatrix}.$$

Clearly

$$\Gamma_L = \Gamma_R A_d,$$

hence

$$A_d = (\Gamma_R^T \Gamma_R)^{-1} \Gamma_R^T \Gamma_L. \tag{7.27}$$

If the pair $(C_2,\ A_d)$ are observable, then $\Gamma_R^T \Gamma_R$ is a full rank matrix and the solution in equation (7.27) is unique. For a 2-input and 2-output system, it is restrictive to assume that $(C_2,\ A_d)$ are observable. But in case where there are many outputs and only one of the outputs is sampled at slow rate, then the assumption that the state can be observed from the fast-sampled outputs is not so restrictive. Pre- and post-multiplying both sides of equation (7.26) by $(\underline{\Gamma_\alpha^d})^\perp$ and $[\frac{1}{J}\underline{U}_{\alpha,J}(k+\beta)\underline{F}_{\alpha,J}^T(k)]^{-1}$ gives

$$\lim_{J\to\infty} \frac{1}{J}(\underline{\Gamma_\alpha^d})^\perp \underline{Y}_{\alpha,J}(k+\beta)\underline{F}_{\alpha,J}^T(k)[\frac{1}{J}\underline{U}_{\alpha,J}(k+\beta)\underline{F}_{\alpha,J}^T(k)]^{-1} = (\underline{\Gamma_\alpha^d})^\perp \underline{H_\alpha^d}. \tag{7.28}$$

After matrices $\underline{\Gamma_\alpha^d}$, C_d, A_d, $\underline{C_d}$ and $\underline{A_d}$ been calculated, it is straightforward to compute the left hand side of equation (7.28); the right hand side of equation (7.28) is a linear combination of matrices B_d and A_d. So matrices B_d and A_d can be computed from equation (7.28) by solving a set of linear equations.

7.6 Noise Effect On Estimation

In earlier sections we assumed that J approaches infinity, but in practice J can only be finite. In this section we will discuss how the noise/disturbance affects the estimation of the process model when J is finite. We will discuss first the open-loop identification case, and then the closed-loop identification case in the latter part of this section.

7.6.1 Open-loop Case

In section 3 we notice that when J is finite, after post-multiplying both sides of equation (7.7) by $\frac{1}{J}U_{\alpha,J}^T(k)$, it follows that

$$\frac{1}{J}Y_{\alpha,J}(k+\beta)U_{\alpha,J}^T(k) = \frac{1}{J}[\Gamma_\alpha^d X_{1,J}^d(k+\beta) + H_\alpha^d U_{\alpha,J}(k+\beta)]U_{\alpha,J}^T(k) + \Xi(J) \quad (7.29)$$

where

$$\Xi(J) = \frac{1}{J}[H_\alpha^s E_{\alpha,J}(k+\beta) + \Gamma_\alpha^s X_{1,J}^s(k+\beta)]U_{\alpha,J}^T(k).$$

Applying the QR decomposition shown in section 7.4 to equation (7.29) and post-multiplying both sides of equation (7.29) by Q_2^J, give

$$\begin{aligned} R_{22}^J &= \frac{1}{J}\Gamma_\alpha^d X_{1,J}^d(k+\beta)U_{\alpha,J}^T(k)Q_2^J + \Xi(J)Q_2^J \quad (7.30)\\ &= \frac{1}{J}\Gamma_\alpha^d(zI - A_d)^{-1}B_d U_{1,J}(k+\beta)U_{\alpha,J}^T(k)Q_2^J + \Xi(J)Q_2^J, \end{aligned}$$

because $(Q_1^J)^T Q_2^J = 0$ and $(Q_2^J)^T Q_2^J = I$. The SVD decomposition of R_{22}^J yields

$$R_{22}^J = U_1^J S_1^J (V_1^J)^T + U_2^J S_2^J (V_2^J)^T, \quad (7.31)$$

where S_1^J contains the n largest singular values of R_{22}^J. Define

$$\Phi_J \triangleq \frac{1}{J}(zI - A_d)^{-1}B_d U_{1,J}(k+\beta)U_{\alpha,J}^T(k)Q_2^J.$$

From equations (7.30) and (7.31) it follows that

$$U_1^J S_1^J = \Gamma_\alpha^d \Phi_J V_1^J + \varepsilon(J), \quad (7.32)$$

where

$$\varepsilon(J) = [\Xi(J)Q_2^J - U_2^J S_2^J (V_2^J)^T]V_1^J. \quad (7.33)$$

Since $(V_2^J)^T V_1^J = 0$, equation (7.33) becomes

$$\begin{aligned} \varepsilon(J) &= \Xi(J)Q_2^J V_1^J \\ &= \frac{1}{J}[H_\alpha^s E_{\alpha,J}(k+\beta) + \Gamma_\alpha^s X_{1,J}^s(k+\beta)]U_{\alpha,J}^T(k)Q_2^J V_1^J. \quad (7.34) \end{aligned}$$

Substituting equation (7.34) into equation (7.32) gives

$$U_1^J S_1^J = \frac{1}{J}\{\Gamma_\alpha^d (zI - A_d)^{-1} B_d U_{1,J}(k+\beta) + H_\alpha^s E_{\alpha,J}(k+\beta) + \Gamma_\alpha^s X_{1,J}^s(k+\beta)\} U_{\alpha,J}^T(k) Q_2^J V_1^J. \tag{7.35}$$

From equation (7.34), it is clear that the estimation error $\varepsilon(J)$ is proportional to $\Xi(J)$. Hence we can say that the larger J is, the smaller the estimation error is. Equation (7.35) indicates that the larger the signal to noise ratio is, the relatively small the estimation error is.

7.6.2 Closed-loop Case

When J is finite, post-multiplying both sides of equation (7.17) by $F_{\alpha,J}^T(k)$ yields

$$\frac{1}{J} Y_{\alpha,J}(k+\beta) F_{\alpha,J}^T(k) = \frac{1}{J}[\Gamma_\alpha^d X_{1,J}^d(k+\beta) F_{\alpha,J}^T(k) + H_\alpha^d U_{\alpha,J}(k+\beta) F_{\alpha,J}^T(k] + \Xi_c(J), \tag{7.36}$$

where

$$\Xi_c(J) = \frac{1}{J}[H_\alpha^s E_{\alpha,J}(k+\beta) + \Gamma_\alpha^s X_{1,J}^s(k+\beta)] F_{\alpha,J}^T(k).$$

After applying the QR decomposition shown in Section 7.4 to $\frac{1}{J} Y_{\alpha,J}(k+\beta) F_{\alpha,J}^T(k)$ and post-multiplying both sides of equation (7.36) by $Q_{2,c}^J$, equation (7.36) becomes

$$R_{22,c}^J = \frac{1}{J}\Gamma_\alpha^d X_{1,J}^d(k+\beta) F_{\alpha,J}^T(k) Q_{2,c}^J + \Xi_c(J) Q_{2,c}^J, \tag{7.37}$$

because $(Q_{1,c}^J)^T Q_{2,c}^J = 0$ and $(Q_{2,c}^J)^T Q_{2,c}^J = I$. Substituting the SVD decomposition of $R_{22,c}$,

$$R_{22,c}^J = U_{1,c}^J S_{1,c}^J (V_{1,c}^J)^T + U_{2,c}^J S_{2,c}^J (V_{2,c}^J)^T, \tag{7.38}$$

where $S_{1,c}^J$ contains the largest n singular values of $R_{22,c}^J$, into equation (7.37), gives

$$U_{1,c}^J S_{1,c}^J = \Gamma_\alpha^d \Lambda_c^d + \varepsilon_c(J),$$

with

$$\begin{aligned}
\Lambda_c^d &= \frac{1}{J} X_{1,J}^d(k+\beta) F_{\alpha,J}^T(k) Q_{2,c}^J V_c^J \\
&= \frac{1}{J}(zI - A_d)^{-1} B_d [(I+GP)^{-1} G F_{1,k} - (I+GP)^{-1} G M E_{1,k}] F_{\alpha,J}^T(k) Q_{2,c}^J V_c^J \\
\varepsilon_c(J) &= [\Xi_c(J) Q_{2,c}^J - U_{2,c}^J S_{2,c}^J (V_{2,c}^J)^T] V_{1,c}^J \\
&= \Xi_c(J) Q_{2,c}^J V_{1,c}^J.
\end{aligned}$$

Again, we can draw a similar conclusion that the larger J is, the smaller the estimation error $\varepsilon_c(J)$ is; and the larger the signal to noise ratio is, the relatively smaller

the estimation error is. $U_1^J S_1^J$ and $U_{1,c}^J S_{1,c}^J$ are the estimated extended observability matrices in the open-loop and closed-loop cases, respectively. When the observation number and the signal to noise ratio are large, the singular values in S_1^J (or $S_{1,c}^J$) are much larger than those in S_2^J (or $S_{2,c}^J$); $U_2^J S_2^J (V_2^J)^T$ (or $U_{2,c}^J S_{2,c}^J (V_{1,c}^J)^T$) mainly contains noise information and $\varepsilon \ll \Gamma_a^d \Phi^J V_1^J$ (or $\varepsilon_c \ll \Gamma_a^d \Lambda_c^d$). So $U_1^J S_1^J$ (or $U_{1,c}^J S_{1,c}^J$) can catch most of the information in the extended observability matrix. When the observation number, J, is finite, the analysis of the noise effect on estimating matrices B_d and D_d is quite involved, and hence is beyond the scope of this thesis.

7.7 Consistency Analysis

If the data length J goes to infinity then the estimated model will converge to the true process model, and such estimation is defined to be consistent. In this section we will discuss the consistency aspects of the proposed algorithm under both open-loop and closed-loop conditions.

7.7.1 Open-loop Analysis

Consistent estimation aspects of some open-loop subspace identification algorithms have been analyzed by Viberg [83], Jansson and Wahlberg [37]. In this sub-section we will analyze the consistency of the open-loop instrumental variable method presented in this chapter. From the derivation of the algorithm, we know that the key to consistent estimation is to estimate the extended observability matrix Γ_a^d consistently. From equation (7.15) we know that $U_1^J S_1^J$ is the estimated extended observability matrix, and

$$\lim_{J \to \infty} U_1^J S_1^J = \Gamma_a^d \Phi_J V_1^J.$$

Define

$$\Lambda^d = \lim_{J \to \infty} \Phi_J V_1^J.$$

Clearly $U_1^J S_1^J$ is the true extended observability matrix if and only if Λ^d is a full rank matrix. From linear algebra theories we know that

$$\det[(\Lambda^d)^T \Lambda^d] = [\det(\Lambda^d)]^2, \tag{7.39}$$

where $\det(\Lambda^d)$ is the determinant of matrix Λ^d. Equation (7.39) implies that if matrix $(\Lambda^d)^T \Lambda^d$ is a full rank matrix so is matrix Λ^d, and vice versa. In the following equations we will prove that given certain conditions $(\Lambda^d)^T \Lambda^d$ is a full rank matrix.

The definition of Λ^d gives

$$\begin{aligned}(\Lambda^d)^T\Lambda^d &= \lim_{J\to\infty}(V_1^J)^T(\Phi_J)^T\Phi_J V_1^J \\ &= \lim_{J\to\infty}\frac{1}{J^2}(V_1^J)^T(Q_2^J)^T U_{\alpha,J}(k)[X_{1,J}^d(k+\beta)]^T X_{1,J}^d(k+\beta)U_{\alpha,J}^T(k)Q_2^J V_1^J.\end{aligned}$$

It is straightforward to get that

$$x_d(k) = (zI - A_d)^{-1}B_d u(k) \tag{7.40}$$

and

$$(zI - A_d)^{-1}B_d = \frac{Q(z^{-1})}{a(z^{-1})},$$

where

$$\begin{aligned}Q(z^{-1}) &= \frac{Adj(zI - A_d)B_d}{z^n} \\ &= Q_1 z^{-1} + Q_2 z^{-2} + \cdots + Q_n z^{-n}\end{aligned} \tag{7.41}$$

and

$$\begin{aligned}a(z^{-1}) &= \frac{\det(zI - A_d)}{z^n} \\ &= 1 + a_1 z^{-1} + \cdots + a_n z^{-n}.\end{aligned}$$

Define

$$\gamma_{xv}(\beta) = \lim_{J\to\infty}\frac{1}{J}X_{1,J}^d(k+\beta)U_{\alpha,J}^T(k) \tag{7.42}$$

Substituting equations (7.40) and (7.41) into equation (7.42) gives

$$\gamma_{xv}(\beta) = \frac{1}{a(z^{-1})}\left[\begin{array}{ccc} Q_n & \cdots & Q_1 \end{array}\right] E\{\left[\begin{array}{c} u(k+\beta-n) \\ \vdots \\ u(k+\beta-1) \end{array}\right]\left[\begin{array}{ccc} u^T(k) & \cdots & u^T(k+\alpha-1) \end{array}\right]\}.$$

For general processes, we assume that the excitation sequence is an ARMA(n_k,n_p) sequence,

$$Q(z^{-1})u(k) = K(z^{-1})v(k),$$

where n_q is the order of $Q(z^{-1})$, n_k is the order of $K(z^{-1})$, and $v(k)$ is a white noise. Then by using Lemma 13 in Jansson and Wahlberg (1998) [37] and condition II of Lemma A3.8 in Soderstrom and Stoica (1983) [69], we can derive the following sufficient conditions for consistent estimation:

1. $(A_d,\ B_d)$ is reachable.
2. $Q(z^{-1})$ and $K(z^{-1})$ are stable.
3. $\max(\beta - \alpha + n + n_k, n + n_q) \leq \alpha$.

7.7.2 Closed-loop Analysis

It is straightforward to show that the estimation is consistent if and only if Λ_c^d is a full rank matrix. From equations (7.24) and (7.23) we have

$$\begin{aligned}(\Lambda_c^d)^T\Lambda_c^d &= \lim_{J\to\infty}[\frac{1}{J}X_{1,J}^d(k+\beta)F_{\alpha,J}^T(k)Q_{2,c}^JV_c^J]^T[\frac{1}{J}X_{1,J}^d(k+\beta)F_{\alpha,J}^T(k)Q_{2,c}^JV_c^J] \\ &= \lim_{J\to\infty}\frac{1}{J^2}[Q_{2,c}^JV_c^J]^TF_{\alpha,J}(k)[X_{1,J}^d(k+\beta)]^TX_{1,J}^d(k+\beta)F_{\alpha,J}^T(k)Q_{2,c}^JV_c^J \quad (7.43)\end{aligned}$$

Define

$$\gamma_{XF}(\beta) = \lim_{J\to\infty}\frac{1}{J}X_{1,J}^d(k+\beta)F_{\alpha,J}^T(k).$$

For general processes, we assume that the excitation sequence is an ARMA(n_k,n_p) sequence,

$$Q(z^{-1})u(k) = K(z^{-1})v(k),$$

and the order of closed-loop system is n_c. From Lemma 13 in Jansson and Wahlberg (1998) [37] and condition II of Lemma A3.8 in Soderstrom and Stoica (1983) [69], it follows that the sufficient conditions for the closed-loop consistent estimation are

1. $(A_d,\ B_d)$ is reachable.
2. $Q(z^{-1})$ and $K(z^{-1})$ are stable.
3. $\max(\beta-\alpha+n+n_k, n+n_c+n_q) \le \alpha$

7.8 Illustrative Examples

A simulation and an experimental example will be presented in this section to demonstrate the proposed closed-loop identification algorithm.

7.8.1 Experimental Examples

The block diagram of a stirred tank heater process is shown in Figure 7.5. This process is a computer-controlled experimental setup at the University of Alberta. The steam through the steam pipe is used to heat the water in the tank; the cold water valve is used to adjust the water level in the tank; and the water outlet valve is fixed. An air-bubble stirrer is used to mix the water in the tank. In this process, the two manipulated variables are the positions of the cold water (u_1) and the steam (u_2) valves; the two measured variables are the water level (y_1) and water temperature (y_2) in the tank. In this experiment, we use signal units of milliampere (mA) to quantify y_1, y_2, u_1 and u_2. The inputs and outputs of the stirred tank heater are summarized below:

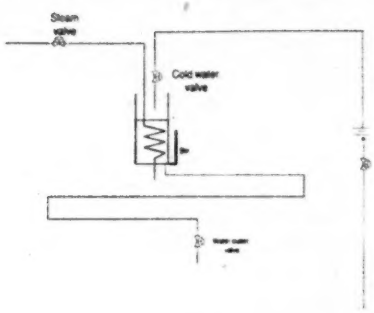


Figure 7.5: *Block diagram of a stirred tank heater system*

Symbol	Quantity	Range
u_1	cold water valve	4 mA $\leq u_1 \leq$ 20 mA
u_2	steam valve	4 mA $\leq u_2 \leq$ 20 mA
y_1	water level	4 mA $\leq y_1 \leq$ 20 mA
y_2	water temperature	4 mA $\leq y_2 \leq$ 20 mA

A single-rate closed-loop experiment and a multirate closed-loop experiment will be performed.

Single-rate Closed-loop Experiment

For this study, the two valve positions were updated every 4 seconds, the water level and water temperature were also sampled every 4 seconds. The controller is a simple PI controller given by:

$$\begin{bmatrix} 1+\frac{0.15}{1-z^{-1}} & 0 \\ 0 & 1.5+\frac{0.1}{1-z^{-1}} \end{bmatrix}$$

We designed Random Binary Sequence (RBS) as the excitation input signal, and added the excitation sequence to the setpoint. The nominal outputs, water level and temperature, were (14 mA, 9 mA). A total of 1566 experimental input-output data points were collected; the first 1000 data points were used to estimate a state-space model with sampling interval of 4 seconds; and the last 566 data points were used for model validation purpose. We then used the controller and the estimated process model to simulate the closed-loop stirred-tank process; the last 566 points of the excitation sequence we used in the experiment were added to the setpoint in the simulation; 566 simulation data points were collected and compared with the last 566 measured experimental data points.

Figures 7.6 and 7.7 show the simulated outputs and the actual measurements for model validation purpose. The values shown in Figures 7.6 and 7.7 are deviation

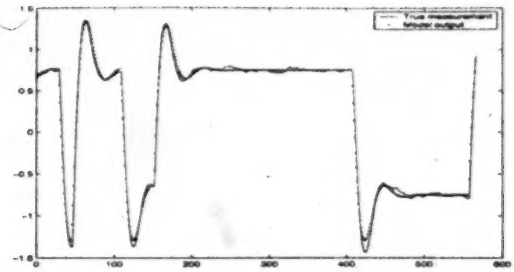


Figure 7.6: *Measured and simulated water levels*

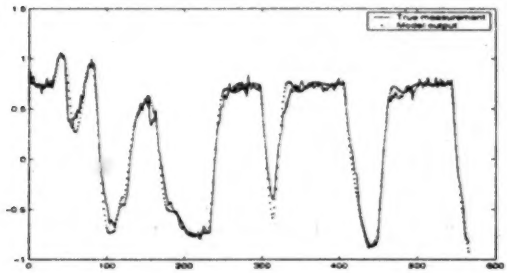


Figure 7.7: *Measured and simulated water temperatures*

values. From the comparison between the simulation and experimental results we can see that the estimated model can catch most of the dynamics of the true process.

Multirate Closed-loop Experiment

In this study, all the control periods are 4 seconds; the temperature is measured every 4 seconds, but the water level is measured every 20 seconds. A multirate inferential PI controller,

$$\begin{bmatrix} 1+\frac{0.15}{1-z^{-1}} & 0 \\ 0 & 1.5+\frac{0.1}{1-z^{-1}} \end{bmatrix},$$

is designed to control the two outputs. The block diagram of the stirred tank heater system with a MR inferential control scheme is shown in Figure 7.8.

In Figure 7.8, $\hat{P}$ represents the estimated fast-sampled model which relates the sampled water level and cold water valve position; y_s represents the estimated water level; r_1 and r_2 are the setpoints; f_1 and f_2 are the excitation sequences. In this experiment, the water level controller G_1 uses both the level measured every 20 seconds and the inter-sample water level predicted by the fast model estimated from the single-rate closed-loop experiment. The excitation, a random binary vector sequence,

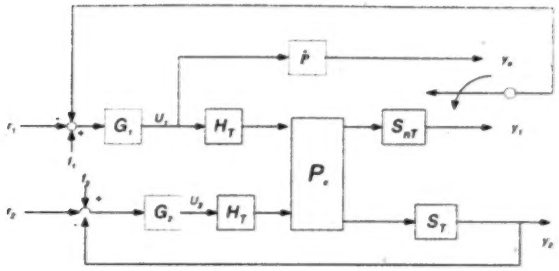


Figure 7.8: *Block diagram of the stirred tank heater system with an inferential control scheme*

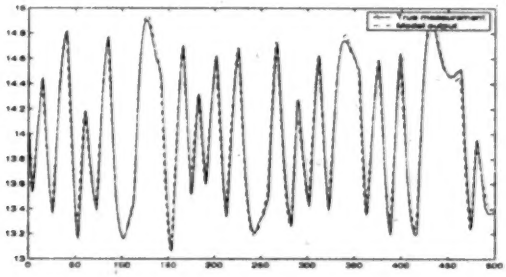


Figure 7.9: *Measured and simulated water level*

is added to the setpoint. The nominal setpoints for water level and temperature were 14 mA and 9 mA respectively. Experimental multirate data over 6.6 hour was collected. The data set over the last 4.4 hours was used to estimate the process model with sampling interval 4 seconds (fast model); the data set over the first 2.2 hours was used for model validation purpose. The estimated process model together with the inferential PI controller simulated the closed-loop stirred-tank system; the excitation sequence over the first 2.2 hours in the experiment was added to the setpoint in the simulation. The model output and the experimental measurement were compared at the fast rates (every 4 seconds). For the sake of clarity, only 500 points (over 33 minutes) are shown in Figures 7.9 and 7.10.

The values shown in Figures 7.9 and 7.10 are deviation values. An unconstrained multirate inferential generalized predictive controller (GPC) [15] in state space framework was designed based on the estimated state-space model of the process. The

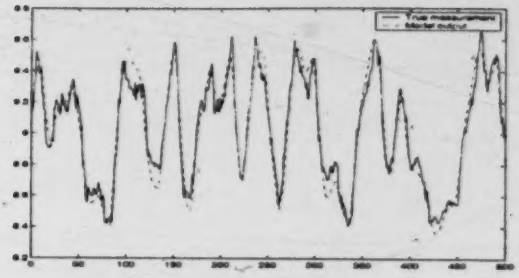


Figure 7.10: *Measured and simulated water temperature*

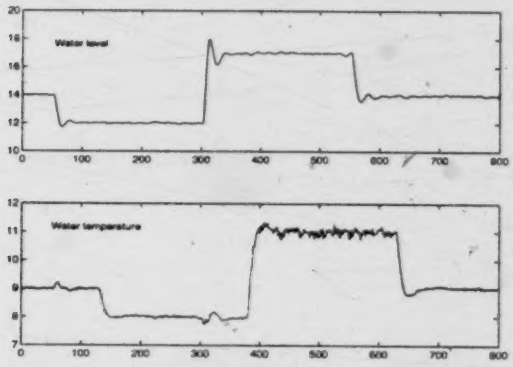


Figure 7.11: *Performance of the stirred tank heater with an MR inferential GPC*

parameters of this GPC are summarized below:

Parameters	Values
Prediction horizon	20
Control horizon	3
Output weighting matrix	Identity matrix
Control weighting matrix	Identity matrix

Several step changes in the setpoint were made, and the output measurements over 3200 seconds are shown in Figure 7.11.

From the performance of the inferential GPC, and the comparison between the simulation and experimental results we can see that the estimated fast-sampled model works well for inferential estimation and control of the true process.

7.8.2 Simulation Examples

A closed-loop system shown in Figure 7.2 is generated as a SIMULINK block diagram. The controller is $G(z) = \frac{0.25}{1-z^{-1}}$; the process model is $\frac{0.5}{5s+1}$ with a unit time delay;

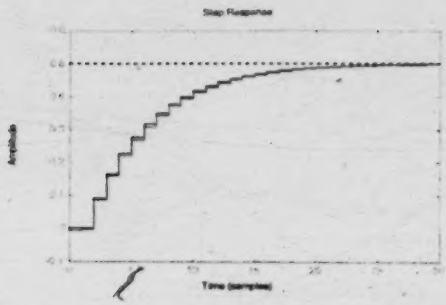


Figure 7.12: *Step response of the true system and the estimated system*

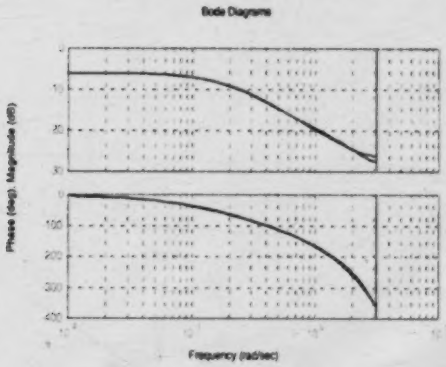


Figure 7.13: *Bode plot of the true system and the estimated system*

and the disturbance is generated by a continuous-time model, $\frac{1}{3s+1}$, with a white noise input e. The sample period is 1 second, and the setpoint is 0. A random binary sequence with 3000 points was generated and added to the control signal as the excitation. The magnitude of the excitation signal was ± 1, and the signal to noise ratio was 2.5 : 1. The control, excitation and output signals were collected, and a process model was estimated by using the proposed closed-loop identification algorithm. The results of the closed-loop identification are shown in Figures 7.12 and 7.13. Figure 7.12 is the step response curves of the true process model and the estimated process model, and Figure 7.13 shows the Bode plots of the true process model and the estimated process model. Figures 7.12 and 7.13 indicate that the estimated model is very close to the true model.

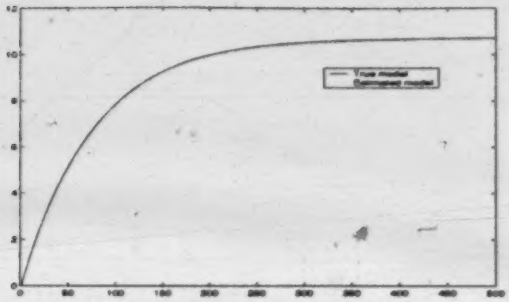


Figure 7.14: *The step responses of the process (relating the water level to the cold water valve) and the estimated model*

Simulated Stirred Tank Heater System

A model,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{0.1416}{z-0.9868} & 0 \\ \frac{-0.0379}{z-0.913} & \frac{0.0595}{z-0.9048} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$

is used to represent the experimental stirred-tank heater process just discussed. A simulation system shown in Figure 7.4 is generated in SIMULINK. The temperature y_2 is sampled every unit of time (4 seconds); the water level y_1 is sampled every 5 units of time (20 seconds); all the holds have uniform period, 1 unit of time. A single-rate PI controller, $G = 1.5 + \frac{0.1}{1-z^{-1}}$, with sampling interval of 1 unit of time (4 seconds), manipulates the steam valve and controls the temperature; a RBS sequence is added to the setpoint of the controller. The water level loop is not closed, and another RBS sequence is used to manipulate the cold water valve. A stochastic process,

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \frac{0.11}{1-0.97z^{-1}} & 0 \\ \frac{-0.02}{1-0.9z^{-1}} & \frac{0.04}{1-0.88z^{-1}} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix},$$

is added to the output of the stirred tank heater process as disturbance. The signal to noise ratio is 5. Simulation data over 5000 units of time was collected and used to estimate the fast model (with period of 4 seconds). The step responses of the true fast model and the estimated fast model are shown in Figures 7.14, 7.15 and 7.16. The figures show that the estimated model can represent the true process very well.

7.9 Conclusion

From the results and discussions in this chapter, it is clear that the proposed instrumental variable method can estimate state-space models of the deterministic systems

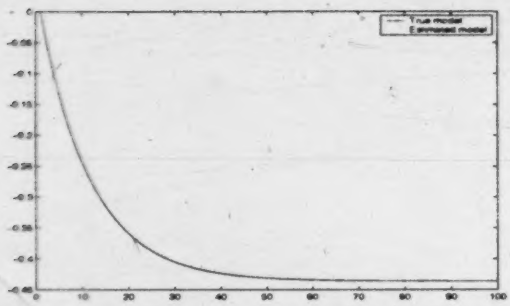


Figure 7.15: *The step responses of the process (relating the water temperature to the cold water valve) and the estimated model*

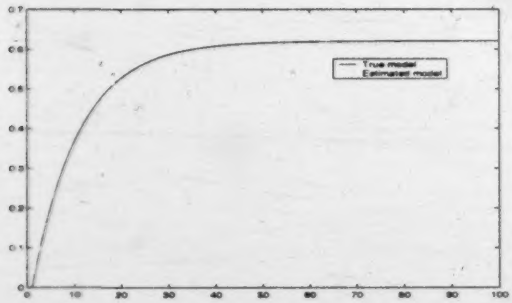


Figure 7.16: *The step responses of the process (relating the water temperature to the steam valve) and the estimated model*

under both open-loop and closed-loop cases even if processes are general BJ type. The proposed instrumental variable methods can also identify the fast-sampled model of a multirate system under closed-loop conditions. From the derivation of the algorithms we can see that this method can be applied to a wide variety of plants, since no specific noise distribution is assumed. Theoretical analysis in this chapter also shows that the proposed identification algorithm can give consistent estimation under both open-loop and closed-loop conditions provided that mild conditions on the system and the input sequence are satisfied.

Chapter 8

Multirate Inferential MPC

8.1 Introduction

Developed towards the end of 1970's, model-based predictive controllers (MPC) [16, 15, 67] capable of handling interactions and constraints have been widely accepted by the process industry. The basic idea in the MPC strategy is to minimize the sum of squares of future control errors over a finite-time horizon and at the same time take into account the incremental control energy required to minimize the control errors. In a receding horizon fashion, model based predictive controllers only implement the first of the calculated control moves.

In Chapter 3 we presented three strategies for controlling multirate systems: slow single-rate control, multirate control with the lifting technique, and fast-rate control with inferential estimation. All the three strategies can be adopted by MPC. Let us use a simple example to illustrate three classes of model-based predictive control schemes for multirate systems. Consider a SISO multirate system where the output sampling interval is nT and the control interval is T. In such a case we can estimate a single-rate model of the process with sampling interval nT and implement a single-rate MPC with period nT. The lifted control strategy can be applied to developing a MPC for this multirate system. The lifted MPC is based on the estimated model of the lifted system: the lifted control moves and outputs of the lifted MPC are inverse lifted to obtain the fast-rate input signals to the process. The block diagram of such a control strategy is shown in Figure 8.1. The basic idea of the inferential MPC is to build a single-rate MPC with interval T which uses both the measured output and the inter-sample outputs predicted by the fast-sampled model.

The major drawback of the slow-rate approach is that it does not fully exploit the capacities of the equipment and the inter-sample performance may be very poor when the least common multiple of the sampling and hold intervals is large. The

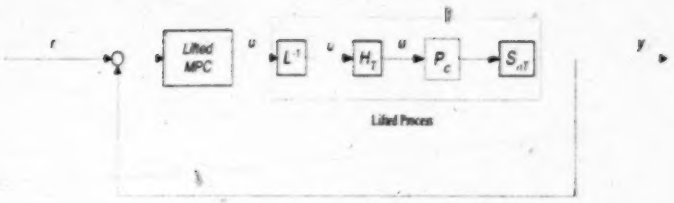


Figure 8.1: *Block diagram of a SISO multirate sampled-data system with a lifted MPC*

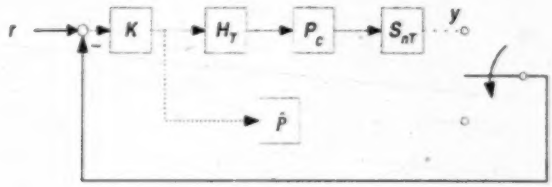


Figure 8.2: *Block diagram of a SISO multirate sampled-data system with an inferential MPC*

difficulties for lifted MPC as shown in Figure 8.2 are the causality constraint and the gain constraint [72]. The gain constraint exists for all kinds of multirate systems; there is no causality constraint for a special class of multirate systems where all the control intervals are uniform, all the output sampling intervals are also uniform, and the output sampling period is an integer multiple of the control interval. So far no methods have been presented to overcome either of the constraints in the lifted MPC framework. If the fast-sampled model is available, then it is straightforward to generate an inferential MPC as shown in Figure 8.2. The only difference between the fast single-rate MPC and the inferential MPC is that the inferential MPC uses the estimated inter-sample outputs but the fast single-rate MPC only uses the measured inter-sample outputs. Kalman-filter-based inferential MPC algorithms are quite popular ([47], [64], [28]), but other statistical-techniques-based inferential MPC algorithms have also been applied successfully ([84], [40]). In this chapter, instead of developing a new multirate inferential MPC, we extend one of the most popular industrial MPC to the multirate inferential control framework.

In this chapter we will only consider a special but practical class of multirate systems where all the hold periods are uniform and all the output sampling intervals are integer multiples of the control interval. This chapter is organized as follows. First an inferential MPC is developed in section 2 for processes where the outputs are corrupted by integrated white noises. Section 3 presents an inferential MPC for another class of disturbances. Experimental evaluation of the proposed inferential

MPC is presented in the last section.

8.2 Inferential MPC

In this section we will focus on one specific class of processes, namely, processes where the process output is corrupted by an additive integrated white noise. When we implement the MPC to dual-rate systems, the output sampling period nT limits the performance of MPC. Knowing the model for the system sampled with period T, we can then use this fast sampled model to estimate the inter-sample outputs $[y(nkT+T), y(nkT+2T), ..., y(nkT+(n-1)T)]$, and design an MPC based on this fast sampled model using both the measured outputs at sampling instant nT and the estimated outputs at the inter-sample instants. This is the basic idea of the model based-inferential control algorithm.

For simplicity, $y(k)$ represents $y(kT)$ in the rest of this chapter. The model of the process sampled with period T is

$$y(k) = P_T(q^{-1})u(k) + \frac{e(k)}{1-q^{-1}}, \tag{8.1}$$

where q^{-1} is a backward shift operator, $P_T(q^{-1})$ is a strictly proper transfer function; $u(k)$ is the control signal; $y(k)$ is the sampled output; $e(k)$ is a white noise input; $u(k)$, $y(k)$ and $e(k)$ are discrete-time signals defined on the time set $\mathcal{Z}_+ := \{0, 1, 2, \cdots\}$ with underlying period T. In the rest of this section, we will develop a dual-rate inferential control algorithm for the system in equation (8.1) in both the transfer function domain and the state space domain.

8.2.1 Inferential Control Algorithm in the Transfer Function Domain

Consider the system in equation (8.1) at time instant nkT. The future output can be expressed as:

$$\begin{aligned} y(nk+j) &= P_T(q^{-1})u(nk+j) + (1+q^{-1}+q^{-2}+...)e(nk+j) \\ &= P_T(q^{-1})u(nk+j) + e(nk+j) + ... + e(nk+1) + \frac{1}{1-q^{-1}}e(nk). \end{aligned} \tag{8.2}$$

Since

$$e(nk) = (1-q^{-1})[y(nk) - P_T(q^{-1})u(nk)], \tag{8.3}$$

substituting equation (8.3) into (8.2) gives

$$y(nk+j) = P_T(q^{-1})u(nk+j) + e(nk+j) + ... + e(nk+1) + y(nk) - P_T(q^{-1})u(nk).$$

Because $e(nk)$ is a white noise signal, the minimum variance prediction of $y(nk+j)$ at time nkT is

$$\hat{y}(nk+j|nk) = P_T(q^{-1})u(nk+j) + y(nk) - P_T(q^{-1})u(nk). \tag{8.4}$$

Rewrite (8.4) as

$$\hat{y}(nk+j|nk) = \frac{P_T(q^{-1})}{\Delta}\Delta u(nk+j) + y(nk) - \frac{P_T(q^{-1})}{\Delta}\Delta u(nk). \tag{8.5}$$

where $\Delta = 1 - q^{-1}$. Let

$$\frac{P_T(q^{-1})}{\Delta} = a_1 q^{-1} + a_2 q^{-2} + ..., \tag{8.6}$$

Substituting (8.6) into (8.5) yields

$$\begin{aligned}
\hat{y}(nk+j|nk) &= a_1\Delta u(nk+j-1) + ... + a_j\Delta u(nk) + ... + y(nk) \\
&\quad - a_1\Delta u(nk-1) - a_2\Delta u(nk-2) - ... \\
&= \sum_{i=1}^{j} a_i\Delta u(nk+j-i) + y(nk) + \sum_{i=1}^{\infty}(a_{i+j} - a_i)\Delta u(nk-i) \qquad (8.7)
\end{aligned}$$

The first term $[\sum_{i=1}^{j} a_i\Delta u(nk+j-i)]$ is the effect of future inputs, and the last term $[y(nk) + \sum_{i=1}^{\infty}(a_{i+j} - a_i)\Delta u(nk-i)]$ represents the expression that defines the free response. We know that when i is large enough, $a_{i+j} \approx a_i \approx K_g$ (K_g is the steady-state gain of the system), so assume that $a_i \approx K_g$ when $i \geq N$ (where N is a sufficiently large integer). Equation (8.7) can then be approximated as

$$\begin{aligned}
\hat{y}(nk+j|nk) &\approx \sum_{i=1}^{j} a_i\Delta u(nk+j-i) + y(nk) + \sum_{i=1}^{N}(a_{i+j} - a_i)\Delta u(nk-i) \qquad (8.8) \\
&\approx \sum_{i=1}^{j} a_i\Delta u(nk+j-i) + \hat{f}(nk+j|nk),
\end{aligned}$$

where

$$\hat{f}(nk+j|nk) \triangleq y(nk) + \sum_{i=1}^{N}(a_{i+j} - a_i)\Delta u(nk-i)$$

denotes the output at time $(nk+j)T$ due to the free response starts from time nkT. Assume that the control action u is constant after time instant $(nk+m-1)T$, i.e.,

$$u(nk+m-1) = u(nk+m) = ... = u(nk+h-1),$$

where h and m are the so called prediction horizon and control horizon, respectively. Now consider the prediction of the output trajectories over the intervals $(nk+1)T$ to $(nk+h)T$,

$$\begin{aligned}
\hat{y}(nk+1|nk) &= \hat{f}(nk+1|nk) + a_1\Delta u(nk) \\
&\vdots \\
\hat{y}(nk+h|nk) &= \hat{f}(nk+h|nk) + a_h\Delta u(nk) + ... + a_{h-m+1}\Delta u(nk+m-1).
\end{aligned}$$

The above equations can be written into a compact vector form as

$$\hat{Y} = \hat{F} + A\Delta U,$$

where

$$\hat{Y} = \begin{bmatrix} \hat{y}(nk+1|nk) \\ \vdots \\ \hat{y}(nk+h|nk) \end{bmatrix}, \ \hat{F} = \begin{bmatrix} \hat{f}(nk+1|nk) \\ \vdots \\ \hat{f}(nk+h|nk) \end{bmatrix}, \ \Delta U = \begin{bmatrix} \Delta u(nk) \\ \vdots \\ \Delta u(nk+m-1) \end{bmatrix},$$

and

$$A = \begin{bmatrix} a_1 & 0 & 0 & ... & 0 \\ a_2 & a_1 & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ a_m & a_{m-1} & a_{m-2} & ... & a_1 \\ ... & ... & ... & ... & ... \\ a_h & a_{h-1} & a_{h-2} & ... & a_{h-m+1} \end{bmatrix}.$$

A set of control actions can be calculated by minimizing the following quadratic objective function

$$J(\Delta U) = (\bar{r} - \hat{Y})^T Q(\bar{r} - \hat{Y}) + \Delta U^T R \Delta U, \tag{8.9}$$

with the constraint

$$\hat{Y} = \hat{F} + A\Delta U,$$

where Q and R are symmetric output weighting and control weighting matrices, respectively; and $\bar{r}$ is a vector containing the h desired future outputs. The solution of this optimization problem is

$$\begin{aligned}
\Delta U &= (A^T Q A + R)^{-1} A^T Q^T (\bar{r} - \hat{F}) \triangleq K_c(\bar{r} - \hat{F}), \\
K_c &= (A^T Q A + R)^{-1} A^T Q^T.
\end{aligned}$$

Since the MPC is a receding horizon based algorithm, only the first control action $\Delta u(nk)$ is implemented. This MPC algorithm which was first developed by Cutler and Ramaker [16] is also know as dynamic matrix control.

From the MPC algorithm we can see that the output is needed every time the control action is calculated. But in the dual-rate system the output is measured only every sample period of nT; if we want to implement MPC every period T, one option is to estimate the $(n-1)$ inter-sample outputs between two successive output measurements. The minimum variance estimation of the $(n-1)$ inter-sample outputs from time nkT to $(n+1)kT$ can be derived from equation (8.8) as follows:

$$\begin{aligned}
\hat{y}(nk+1) &= y(nk) + \sum_{i=1}^{N}(a_{i+1} - a_i)\Delta u(nk-i) + a_1\Delta u(nk) \\
&\vdots \\
\hat{y}(nk+n-1) &= y(nk) + \sum_{i=1}^{N}(a_{i+n-1} - a_i)\Delta u(nk-i) \\
&\quad + \sum_{i=0}^{n-2} a_{n-1-i}\Delta u(nk+i).
\end{aligned}$$

The unconstrained inferential MPC algorithm then consists of the following steps:

Step 1 Calculate $K_c = (A^TQA + R)^{-1}A^TQ^T$.

Step 2 Calculate the prediction of the output trajectories over the prediction horizon of interest. At time instant knT the output is measured, and the free response can be calculated as follows,

$$\hat{f}(nk+j|nk) = y(nk) + \sum_{i=1}^{N}(a_{i+j} - a_i)\Delta u(nk-i) \quad j = 1,..,h.$$

At time instant $t = knT + lT$, $1 \le l < n$, the output is not measured; but we can estimate the inter-sample output

$$\begin{aligned}
\hat{y}(nk+l) &= y(nk) + \sum_{i=1}^{N}(a_{i+l} - a_i)\Delta u(nk-i) + \\
&\quad a_l\Delta u(nk+1) + ... + a_1\Delta u(nk+l-1);
\end{aligned}$$

and then compute the free response based on $\hat{y}(nk+i)$,

$$\hat{f}(nk+l+j|nk+l) = \hat{y}(nk+l) + \sum_{i=1}^{N}(a_{i+j} - a_i)\Delta u(nk+l-i),$$

where $j = 1, 2, ..., h$.

Step 3 Calculate only the next control action

$$\Delta u(t) = k_1(\bar{r} - \hat{f}),$$

where k_1 is the first row of the controller matrix K_c.

8.2.2 Inferential Control Algorithm in the State-Space Domain

The system in equation (8.1) can be represented as

$$\begin{aligned} y(k) &= P_T(q^{-1})\frac{1}{\Delta}[\Delta u(k)] + \frac{1}{\Delta}e(k) \\ &= y_d(k) + y_s(k), \end{aligned} \tag{8.10}$$

where

$$\begin{aligned} y_d(k) &= [P_T(q^{-1})\frac{1}{\Delta}][\Delta u(k)] \\ y_s(k) &= \frac{1}{1-q^{-1}}e(k). \end{aligned}$$

$y_d(k)$ and $y_s(k)$ are the outputs of the deterministic and stochastic parts of the process, respectively. The state-space model of $P_T(q^{-1})\frac{1}{\Delta}$ can be easily computed from the state-space model of $P_T(q^{-1})$. Assume $P_T(q^{-1})\frac{1}{\Delta}$ can be realized by the following state-space model,

$$\begin{aligned} x_d(k+1) &= A_I x_d(k) + B_I \Delta u(k) \\ y_d(k) &= C_I x_d(k). \end{aligned}$$

Consider the prediction of the h future outputs at time nkT, and define

$$\hat{Y}(nk) = \begin{bmatrix} \hat{y}(nk+1|nk) \\ \vdots \\ \hat{y}(nk+h|nk) \end{bmatrix}.$$

The predictions can then be expressed as

$$\hat{Y}(nk) = \hat{Y}_d(nk) + \hat{Y}_s(nk),$$

where

$$\hat{Y}_d(nk) = \begin{bmatrix} \hat{y}_d(nk+1|nk) \\ \vdots \\ \hat{y}_d(nk+h|nk) \end{bmatrix}, \quad \hat{Y}_s(nk) = \begin{bmatrix} \hat{y}_s(nk+1|nk) \\ \vdots \\ \hat{y}_s(nk+h|nk) \end{bmatrix}.$$

It is straightforward to derive the following expression for $\hat{Y}_d(nk)$,

$$\hat{Y}_d(nk) = \begin{bmatrix} C_I A_I \\ C_I A_I^2 \\ \vdots \\ C_I A_I^h \end{bmatrix} x_d(nk) + \begin{bmatrix} C_I B_I & 0 & \cdots & 0 \\ C_I A_I B_I & C_I B_I & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ C_I A_I^{h-1} B_I & C_I A_I^{h-2} B_I & \cdots & C_I B_I \end{bmatrix} \begin{bmatrix} \Delta u(nk) \\ \Delta u(nk+1) \\ \vdots \\ \Delta u(nk+h-1) \end{bmatrix}. \tag{8.11}$$

When the control horizon is m,

$$\Delta u(nk+m) = \Delta u(nk+m+1) = \cdots = \Delta u(nk+n-1) = 0;$$

in which case equation (8.11) becomes

$$\hat{Y}_d(nk) = \Gamma x_d(nk) + H\Delta U, \tag{8.12}$$

where

$$\Gamma = \begin{bmatrix} C_I A_I \\ C_I A_I^2 \\ \vdots \\ C_I A_I^h \end{bmatrix}, \quad H = \begin{bmatrix} C_I B_I & 0 & \cdots & 0 \\ C_I A_I B_I & CB & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ C_I A_I^{h-1} B_I & C_I A_I^{h-2} B_I & \cdots & C_I A_I^{h-m} B_I \end{bmatrix}, \quad \Delta U = \begin{bmatrix} \Delta u(nk) \\ \Delta u(nk+1) \\ \vdots \\ \Delta u(nk+m-1) \end{bmatrix}$$

and H is the so called step response coefficient matrix. It is true that

$$y(nk) = C_I x_d(nk) + y_s(nk),$$

so

$$y_s(nk) = y(nk) - C_I x_d(nk).$$

Since the disturbance is an integrated white noise,

$$\begin{aligned} y_s(nk+j) &= y_s(nk) + e(nk+1) + \cdots + e(nk+j) \\ &= y(nk) - C_I x_d(nk) + e(nk+1) + \cdots + e(nk+j) \end{aligned} \tag{8.13}$$

Combining equations (8.12) and (8.13) gives the minimum variance prediction of $\hat{Y}(nk)$ at time instant nkT ,

$$\hat{Y}(nk) = \Gamma x_d(nk) + H\Delta U(nk) + \hat{Y}_s(nk),$$

where

$$\hat{Y}_s(nk) = \begin{bmatrix} y(nk) - C_I x_d(nk) \\ \vdots \\ y(nk) - C_I x_d(nk) \end{bmatrix} \tag{8.14}$$

The solution of the minimization problem (8.9) is

$$\Delta U(nk) = (H^T Q H + R)^{-1} H^T Q^T [\bar{r} - \Gamma x_d(nk) - \hat{Y}_s(nk)].$$

where $\bar{r}$, Q and R are as before. In the dual-rate system case, the output is measured only every nT, but we can estimate the $(n-1)$ inter-sample outputs as follows:

$$\begin{aligned} \hat{y}_d(nk+1|nk) &= C_I A_I x_d(nk) + C_I B_I \Delta u(nk) + y(nk) - C_I x_d(nk) \\ &\vdots \\ \hat{y}_d(nk+n-1|nk) &= C_I A_I^{n-1} x_d(nk) + \sum_{i=1}^{n-1} C_I A_I^{i-1} B_I \Delta u(nk+i-1) + y(nk) - C_I x_d(nk). \end{aligned}$$

After estimating the inter-sample outputs, we can implement the dual-rate inferential MPC in a straightforward way.

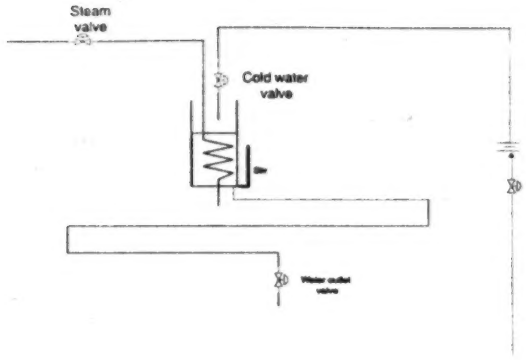


Figure 8.3: *Experimental setup of a stirred tank heater*

8.3 Experimental Evaluation

The MIMO process shown in Figure 8.3 is a computer-controlled experimental setup at the University of Alberta and is concerned with temperature and water level regulation in a continuously stirred tank heater.

In this process, the two manipulated variables are the positions of the cold water (u_1) and the steam (u_2) valves; the two measured variables are the water level (y_1) and water temperature (y_2) in the tank. For this study, the two valve positions are updated every 4 second, the water level and water temperature are sampled every 20 second and 4 second, respectively. Around the operating point with $y_1 = 12$ mA and $y_2 = 10$ mA, the inputs and outputs of the stirred tank heater are summarized below:

Symbol	Quantity	Range
u_1	cold water valve	$4 \text{ mA} \leq u_1 \leq 20 \text{ mA}$
u_2	steam valve	$4 \text{ mA} \leq u_2 \leq 20 \text{ mA}$
y_1	water level	$4 \text{ mA} \leq y_1 \leq 20 \text{ mA}$
y_2	water temperature	$4 \text{ mA} \leq y_2 \leq 20 \text{ mA}$

We use mA to quantify both y_1 and y_2 since there are simple linear relationships to translate these units to actual physical units. From the multirate input-output data a single-rate model with sampling interval of 4 seconds was identified(details in chapter 3). We changed the water level sampling period to 40 seconds and kept the other sampling rates to 4 seconds. Based on this fast sampled model, we designed an inferential MPC and applied it to the multirate system. The parameters of this

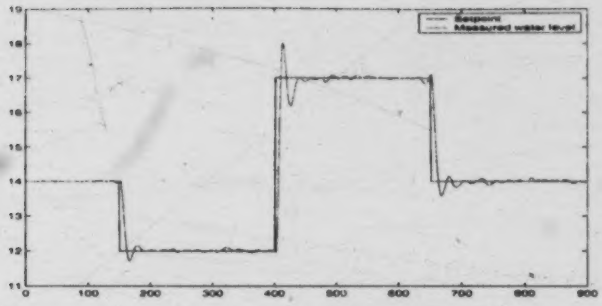


Figure 8.4: *The measured water level*

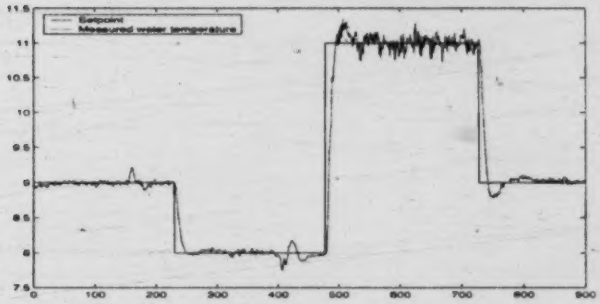


Figure 8.5: *The measured water temperature*

inferential MPC are summarized below:

Parameters	Values
Prediction horizon	12
Control horizon	3
Output weighting matrix	Identity matrix
Control weighting matrix	Identity matrix

We made several step changes in the setpoint, Figures 8.4 and 8.5 show trajectories of the outputs measured every 4 seconds (the inferential controllers uses the water level sampled every 40 seconds).

There is some noise in the temperature measurement, that is because the stirrer, which makes the water in the tank well mixed, was broken when we did this experiment. Some ripples in the measured water level and water temperature are due to interaction in the system.

8.4 Conclusion

In this chapter we used both theoretical analysis and experimental results to show that the multirate inferential control algorithms work well in a model-based predictive control framework.

Chapter 9

Analysis of Multirate Inferential Systems

9.1 Introduction

Instability of the feedback system has been one of the most serious problems after the first automatic feedback controller was used in the industry. This in turn prompted mathematical analysis of feedback systems; Maxwell (1868) and Vyshnegradskii (1877) were pioneers in analyzing the stability of feedback control systems. The observed stability problems of the negative feedback control systems were explained through the frequency domain analysis techniques developed by Nyquist [62] and Bode [5] and Black [4]. Youla and co-workers [87, 88] showed that all stabilizing controllers for a particular system can be parametrized in an effective manner. The Youla parametrization simplifies the task of searching for a good stabilizing controller dramatically. Brosilow [7] used the Internal Model Control structure to explain the special case of Youla parametrization. In practice, no model is a perfect representation of the process, so model-plant mismatch is inevitable. The robust stability (i.e., stability of the feedback systems in the presence of model uncertainty) and robust performance (i.e., performance of the feedback systems in the presence of model uncertainty) have attracted much attention because of their practical importance. Doyle developed a powerful tool (the structured singular value) for testing robust stability and robust performance [18, 19].

Multirate inferential control algorithms have been applied to chemical industrial processes. Intuitively, the nominal performance and stability of the multirate system with an inferential controller should be close to those of the single-rate system at the fast rate when the model is a perfect representation of the fast-sampled process. However model-plant mismatch (MPM) is inevitable in the chemical process industry

due to factors such as scaling, fouling, varying catalyst activity (in chemical reactors), equipment degradation with time, etc. So it is important to know the stability of the dual-rate inferential system in the presence of model-plant mismatch. No work has been reported on robust performance and stability of the multirate inferential systems. For simplicity we focus here on a special but practical class of dual-rate systems where the output sampling interval is an integer multiple of the control interval. The main objective of this chapter is to develop a framework for the performance and stability analysis of the dual-rate system with an inferential estimation algorithm for feedback.

This chapter is organized as follows. In Section 2 we briefly introduce lifting techniques and lifted systems in the transfer function framework. The dual-rate inferential control scheme is introduced in Section 3. The performance of the inferential control scheme in the absence of model-plant mismatch is visited in Section 4. Section 5 discusses stability robustness of inferential control systems in the presence of model-plant mismatch. In Section 6 we give an illustrative example to demonstrate the results given in Section 5. Finally we discuss some extension to the result given in Section 5 followed by the concluding marks.

9.2 Preliminaries

Assume that the transfer function $P_T(z)$ represents the discrete model of process P with sampling interval T; we know that $P_T(z)$ can be expressed as:

$$P_T(z) = p(0) + z^{-1}p(1) + z^{-2}p(2) + ...,$$

where z is the forward shift operator. Using polyphase decomposition [74], we define

$$\begin{aligned}
P_0(z^n) &\triangleq p(0) + z^{-n}p(n) + z^{-2n}p(2n) + ... \\
P_1(z^n) &\triangleq p(1) + z^{-n}p(n+1) + z^{-2n}p(2n+1) + ... \\
&\cdots \\
P_{n-1}(z^n) &\triangleq p(n-1) + z^{-n}p(2n-1) + z^{-2n}p(3n-1) + ...
\end{aligned}$$

Then $P_T(z)$ can be written as

$$P_T(z) = P_0(z^n) + z^{-1}P_1(z^n) + ... + z^{1-n}P_{n-1}(z^n).$$

Define

$$\underline{P} = L_n P_T L_n^{-1}.$$

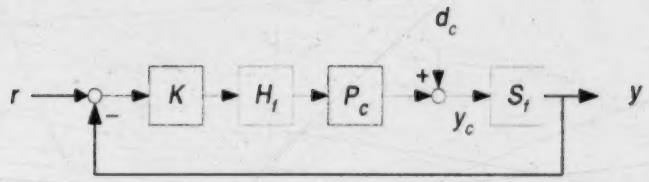


Figure 9.1: *The sampled-data single-rate control system.*

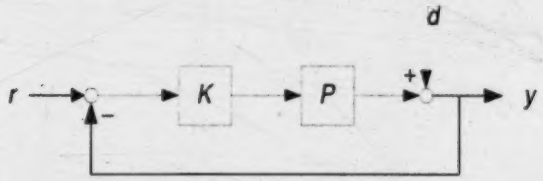


Figure 9.2: *The discrete-time single-rate control system.*

The lifted system $\underline{P}$ can be expressed as [42]

$$\underline{P}(z) = \begin{bmatrix} P_0(z) & z^{-1}P_{n-1}(z) & \dots & z^{-1}P_1(z) \\ P_1(z) & P_0(z) & \dots & z^{-1}P_2(z) \\ \dots & \dots & \dots & \dots \\ P_{n-1}(z) & P_{n-2}(z) & \dots & P_0(z) \end{bmatrix}.$$

9.3 Dual-Rate Inferential Control Scheme

First, let us consider a single-input, single-output single-rate control system shown in Figure 9.1, where P_c is a continuous-time LTI plant and K a digital controller. The two systems P_c and K are interfaced by the A/D and D/A converters, modeled by S_f, the ideal sampler, and H_f, the zero-order hold (ZOH), respectively, both operating with the *fast* period T. This is a single-rate sampled-data control system which involves two exogenous signals, the discrete-time reference $r(k)$ and the continuous-time disturbance $d_c(t)$. The measured continuous-time output is $y_c(t)$. Define P as the ZOH equivalent model of P_c ($P = S_f P_c H_f$) and discretize $d_c(t)$ at the fast rate: $d(k) = d_c(kT)$. Thus Figure 9.1 is equivalent to a pure discrete-time control system in Figure 9.2, which involves only discrete-time signals [10].

Suppose that due to physical constraints, we cannot sample the output as fast as we wish and thus we have to replace S_f in Figure 9.1 by a *slow* sampler S_s with a sampling period nT, where n is a positive integer: $n > 1$. In order to maintain single-rate control, one option is to adopt a slow zero-order hold H_s with period

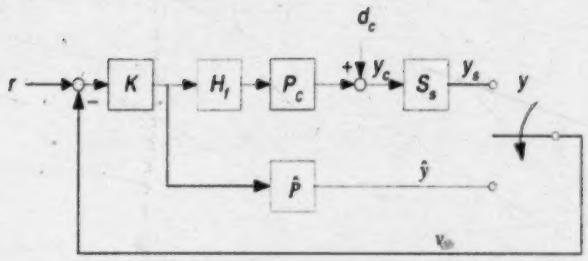


Figure 9.3: *The sampled-data inferential control system.*

nT and obtain a single-rate control system operating at the slow rate; however, the disadvantage is that performance degradation could be significant. The option that we propose is the inferential control scheme shown in Figure 9.3, where the output sampling is now slow (S_s), but the fast zero-order hold H_f and the fast single-rate controller K are still in place. For such a scheme to work, we assume that a model $\hat{P}$ for the fast single-rate system P is available. In order to feed back to the controller K a fast rate signal $v(k)$, representing the output $y(k)$, we use the slow sampled output ($y_s(k) = y_c[k(nT)]$) every nT period, giving $y(0)$, $y(n)$, and $y(2n)$, etc., and use the model $\hat{P}$ to get the estimated output $\hat{y}(k)$ to fill in the missing samples in $y(k)$. Such a process is depicted in Figure 9.3 by a periodic switch which connects to y_s at times $t = j(nT)$, and connects to $\hat{y}(k)$ at $t = j(nT) + iT$, $i = 1, 2, \cdots, n-1$. Thus the output of the switch is a fast rate signal given by

$$v(k) = \begin{cases} y_s(j), & k = jn, \\ \hat{y}(jn+i), & k = jn+i, 0 < i < n. \end{cases}$$

Since S_s is the same as S_f followed by the periodic switch shown in Figure 9.3, it is easy to see that the equivalent discrete-time model for Figure 9.3 is Figure 9.4. Here, P, d and y are as before. Due to the periodic switch, the fictitious fast rate signal $y(k)$ is fed back only once every n samples. Therefore

$$v(k) = \begin{cases} y(jn), & k = jn, \\ \hat{y}(jn+i), & k = jn+i, 0 < i < n. \end{cases}$$

To summarize, the dual-rate inferential control scheme uses a fast-rate plant model, a fast single-rate controller, and a periodic switch. It is conceptually simple, easy to implement in digital computers, and practical for industry. Later we will show that in comparison with the fast single-rate control system in Figure 9.1, we may lose some performance; but we will gain robustness.

Note that the inferential control scheme assumes availability of a fast single-rate model $\hat{P}$. There are two ways of obtaining $\hat{P}$: (1) if a model for the continuous-time

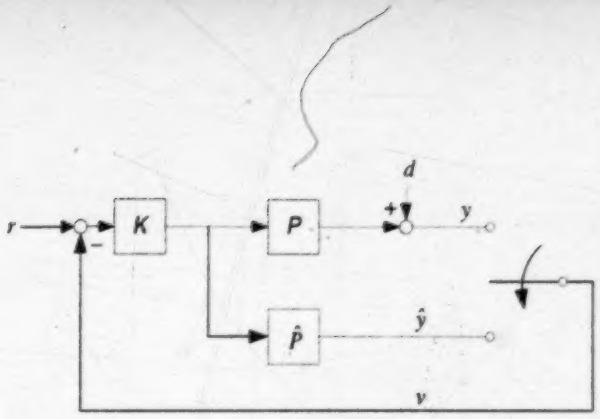


Figure 9.4: *The discrete-time inferential control system.*

plant P is available, $\hat{P}$ can be computed easily by discretization; (2) if not, we need to invoke the results in chapter 4 to identify such a fast model based on multirate input-output data.

Next we will compare the inferential control system in Figure 9.4 with the single-rate system in Figure 9.2 in tracking and disturbance rejection performance (Section 4) and in stability robustness (Section 5). Conclusions are given in the last section.

9.4 Nominal Performance

Consider the dual-rate inferential control system in Figure 9.4. Assume in this section that there is no MPM; thus $\hat{P} = P$. If furthermore there is no disturbance in the system, i.e., $d(k) = 0$, then $\hat{y}(k) = y(k)$ and hence $v(k) \equiv y(k)$. Thus the dual-rate system is equivalent to the single-rate system in Figure 9.2. We thus conclude that:

- Without MPM, closed-loop stability of the dual-rate system in Figure 9.4 is equivalent to that of the single-rate system in Figure 9.2.
- Without MPM and disturbances, the tracking performance (y following r) of Figure 9.4 is the same as that of Figure 9.2.

This is the main reason why the proposed dual-rate inferential control scheme is attractive: In the ideal situation, we can expect to recover the performance of the fast single-rate system.

We now examine the disturbance rejection capability of the two system involved when a disturbance is present ($d \neq 0$). First, let us look at the single-rate system in Figure 9.2. Defining the system from d to y as G_{sr}, we get

$$G_{sr} = (I + PK)^{-1}. \tag{9.1}$$

We can use the $\mathcal{H}_\infty$ norm to quantify the effect of d on y as follows: Suppose the significant frequency components in d are captured by the pass-band of a pre-filter W_2; the quantity $\|G_{sr}W_2\|_\infty$ is then appropriate as a worst-case measure of the effect of disturbance. The best achievable disturbance rejection performance, denoted γ_{sr}, is obtained by minimizing $\|G_{sr}W_2\|_\infty$ over the class of controllers providing closed-loop stability – a standard $\mathcal{H}_\infty$ optimization problem. In the special case when P is already stable, we can parametrize the set of stabilizing controllers via

$$K = (I - QP)^{-1}Q \tag{9.2}$$

with Q stable and LTI. Substituting 9.2 into 9.1 gives

$$\begin{aligned}(I + PK)^{-1} &= (I + P(I - QP)^{-1}Q)^{-1} \\ &= I - PQ,\end{aligned}$$

since

$$\begin{aligned}&(I + P(I - QP)^{-1}Q)(I - PQ) \\ =\ & I - PQ + P(I - QP)^{-1}(I - QP)Q \\ =\ & I.\end{aligned}$$

Thus we arrive at the following model-matching problem:

$$\begin{aligned}\gamma_{sr} &= \min_Q \|G_{sr}W_2\|_\infty \\ &= \min_Q \|(I - PQ)W_2\|_\infty.\end{aligned} \tag{9.3}$$

The minimization is done over the class of stable and LTI Q's. The quantity γ_{sr} can be thought of as a measure of disturbance rejection capability of the single-rate system in Figure 9.2.

Next, we look at the dual-rate system in Figure 9.4. Define G_{dr} as the closed-loop system from d to y in Figure 9.4. Thus

$$\gamma_{dr} = \min \|G_{dr}W_2\|_\infty$$

represents the disturbance rejection capability of the dual-rate system in Figure 9.4.

Because of the presence of the switch, the system in Figure 9.4 is linear and periodically time-varying. In order to derive a model for G_{dr}, we use the standard lifting technique [41, 10].

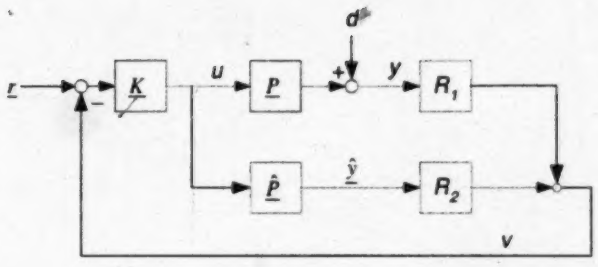


Figure 9.5: *The lifted inferential control system.*

Let $x(k)$ $(0 \leq k < \infty)$ be a discrete-time signal. The lifted signal $\underline{x}$ is defined as

$$\underline{x} = \left\{ \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(n-1) \end{bmatrix}, \begin{bmatrix} x(n) \\ x(n+1) \\ \vdots \\ x(2n-1) \end{bmatrix}, \cdots \right\}.$$

Note that after lifting, the signal dimension is increased by a factor of n, as is the underlying period.

Lifting all the signals involved in Figure 9.4 to get $\underline{y}$ for y, $\underline{\hat{y}}$ for $\hat{y}$, etc., we can derive a simple model for the periodic switch which relates v to y and $\hat{y}$ in the lifted domain:

$$\underline{v} = R_1\underline{y} + R_2\underline{\hat{y}}. \tag{9.4}$$

Here R_1 and R_2 are static systems given by the following matrices:

$$R_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}_{n\times n}, \quad R_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{n\times n}. \tag{9.5}$$

The lifted systems

$$\underline{K} = LKL^{-1}, \quad \underline{P} = LPL^{-1}, \underline{\hat{P}} = L\hat{P}L^{-1},$$

together with (9.4) give rise to the lifted closed-loop system in Figure 9.5, which is the equivalent model under lifting for the dual-rate structure in Figure 9.4. The advantage is that we are now dealing with an LTI system.

Specializing to our discussion in this section, we set $\underline{r} = 0$ and $\underline{\hat{P}} = \underline{P}$. The lifted system $\underline{G}_{dr} := LG_{dr}L^{-1}$, or equivalently, the system from $\underline{d}$ to $\underline{y}$ in Figure 9.5, can be derived as follows. First, compute the system from $\underline{d}$ to $\underline{u}$ (noting that $R_1 + R_2 = I$):

$$\underline{u} = -(I + \underline{K}\,\underline{P})^{-1}\underline{K}R_1\underline{d}.$$

Then since $\underline{y} = \underline{d} + \underline{P}\,\underline{u}$, we have

$$\underline{G}_{dr} = I - \underline{P}(I + \underline{K}\,\underline{P})^{-1}\underline{K}R_1. \tag{9.6}$$

If P is stable, we can use (9.2) for controller parametrization; the lifted version is

$$\underline{K} = (I - \underline{Q}\,\underline{P})^{-1}\underline{Q}.$$

Substituting this into (9.6) yields:

$$\begin{aligned}
\underline{G}_{dr} &= I - \underline{P}(I + \underline{K}\underline{P})^{-1}\underline{K}R_1 \\
&= I - \underline{P}[I + (I - \underline{Q}\underline{P})^{-1}\underline{Q}\underline{P}]^{-1}(I - \underline{Q}\underline{P})^{-1}\underline{Q}R_1 \\
&= I - \underline{P}(I - \underline{Q}\underline{P})(I - \underline{Q}\underline{P})^{-1}\underline{Q}R_1 \\
&= I - \underline{P}\underline{Q}R_1.
\end{aligned}$$

Since lifting preserves norms, we have

$$\gamma_{dr} = \min \|\underline{G}_{dr}\underline{W}_2\|_\infty = \min \|(I - \underline{P}\,\underline{Q}R_1)\underline{W}_2\|_\infty, \tag{9.7}$$

where the latter minimization is over the class of stable and LTI Q's.

In order to compare γ_{sr} and γ_{dr}, we lift the systems involved in (9.3) to get

$$\gamma_{sr} = \min \|(I - \underline{P}\,\underline{Q})\underline{W}_2\|_\infty.$$

Now suppose Q^* is the optimal solution for the minimization in (9.7), i.e.,

$$\gamma_{dr} = \|(I - \underline{P}\,\underline{Q}^*R_1)\underline{W}_2\|_\infty.$$

(If γ_{dr} is not attainable, we can use a sequence of Q^*s so that the performance converges to γ_{dr}, and the argument to follow is similar.) Define $\underline{Q}_1 = \underline{Q}^*R_1$. The corresponding Q_1 is stable but not LTI; it is in fact linear and periodic with period n (because R_1 does not correspond to an LTI system before lifting). However, since for LTI plants, linear periodic control does not offer any advantage over LTI control [41], we conclude

$$\gamma_{dr} = \|(I - \underline{P}\,\underline{Q}_1)\underline{W}_2\|_\infty \geq \min \|(I - \underline{P}\,\underline{Q})\underline{W}_2\|_\infty = \gamma_{sr}.$$

This result can be summarized as follows:

Proposition 5 *When P is stable and there is no MPM, the disturbance rejection capability of the dual-rate system in Figure 9.4 is no better than that of the single-rate system in Figure 9.2.*

Proposition 5 perhaps makes sense intuitively; but it is not clear if it is still valid when P is unstable. In the next section, we give a somewhat surprising result on stability robustness.

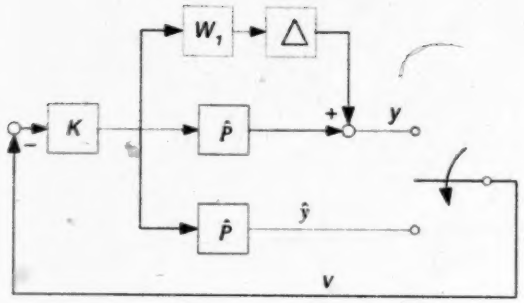


Figure 9.6: *The inferential control system with additive uncertainty.*

9.5 Stability Robustness

In this section we assume that there is MPM in the dual-rate system in Figure 9.4, and hence $\hat{P} \neq P$. We will study issues related to stability robustness. We treat P as uncertain and $\hat{P}$ as the nominal plant model; we assume a standard additive uncertainty model [20], i.e., P belongs to the uncertainty class given by

$$\{\hat{P} + \Delta W_1 : \ \|\Delta\|_\infty < 1\}.$$

The MPM is represented by ΔW_1, where Δ is the perturbation, assumed to be stable and LTI, with norm less than 1 (normalized), and W_1 is a fixed frequency weighting filter which is stable and LTI. The inferential control system with this uncertain structure is depicted in Figure 9.6. Our goal is to find a condition under which the closed-loop system is stable for all admissible Δ.

Proposition 6 *Assume that K stabilizes $\hat{P}$ (nominal stability). Let $\underline{W}_1$, $\underline{K}$, and $\underline{\hat{P}}$ be the lifted systems of W_1, K, and $\hat{P}$, respectively. The dual-rate system in Figure 9.6 is closed-loop stable for all admissible Δ if*

$$\|\underline{W}_1(I + \underline{K}\,\underline{\hat{P}})^{-1}\underline{K}R_1\|_\infty < 1, \tag{9.8}$$

where the matrix R_1 was defined in (9.5).

Proof Similar to what we did in Section 3, we lift the system in Figure 9.6 to get Figure 9.7.

Isolating $\underline{\Delta}$, we can reconfigure Figure 9.7 into Figure 9.8, where $\underline{M}$ is given by

$$\underline{M} = -\underline{W}_1(I + \underline{K}\,\underline{\hat{P}})^{-1}\underline{K}R_1.$$

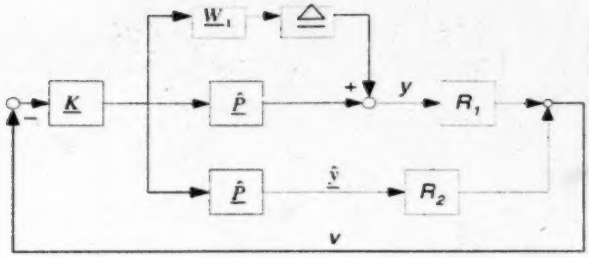


Figure 9.7: *The lifted inferential control system with additive uncertainty.*

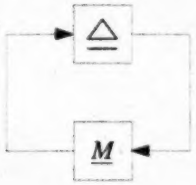


Figure 9.8: *The equivalent system of Figure 9.7.*

It follows easily from the nominal stability assumption that $\underline{M}$ is a stable system. Applying the small-gain condition to the feedback system in Figure 9.8, we conclude that the closed-loop system is stable for all admissible Δ if

$$\|\underline{\Delta}\|_\infty \cdot \|\underline{M}\|_\infty < 1,$$

which is true for all admissible Δ ($\|\underline{\Delta}\|_\infty < 1$) if $\|\underline{M}\|_\infty < 1$. ∎

Now we compare the robustness condition in Proposition 6 with that for the single-rate system of Figure 9.2. Such a condition was well-known [20]: The closed-loop system in Figure 9.2 is stable for all admissible Δ if

$$\|W_1(I + K\hat{P})^{-1}K\|_\infty < 1,$$

which is equivalent to the following after lifting:

$$\|\underline{W}_1(I + \underline{K}\,\hat{\underline{P}})^{-1}\underline{K}\|_\infty < 1.$$

The quantity on the left, denoted β_{sr}, can be used as a measure of stability robustness for the single-rate system [20]: The smaller the β_{sr} is, the more robust the system is. Similarly, the quantity on the left of (9.8), denoted β_{dr}, is a robustness measure for

the dual-rate inferential control system. Examining the two quantities, we get

$$\begin{aligned}\beta_{dr} &= \left\| \underline{W}_1 (I + \underline{K}\hat{\underline{P}})^{-1} \underline{K} R_1 \right\|_\infty \\ &\leq \left\| \underline{W}_1 (I + \underline{K}\hat{\underline{P}})^{-1} \underline{K} \right\|_\infty \|R_1\|_\infty \\ &\leq \left\| \underline{W}_1 (I + \underline{K}\hat{\underline{P}})^{-1} \underline{K} \right\|_\infty \\ &\leq \beta_{sr},\end{aligned}$$

and hence we conclude:

Corollary 1 *The fast single-rate control system is no more robust than the dual-rate inferential control system.*

This result still leaves room for doubt: Can the dual-rate control structure be *better* in stability robustness than the fast single-rate one? The answer is positive; and we illustrate this with an example in the next section.

9.6 Illustrative Example

Example Consider a nominal plant $\hat{P}$ with a PI controller K:

$$\hat{P}(z) = \frac{0.15}{z - 0.9}, \quad K(z) = 3 + \frac{0.5}{1 - z^{-1}}.$$

Suppose the actual plant P, different from $\hat{P}$, is given by

$$P(z) = \frac{0.13}{z^3(z - 0.92)}.$$

Clearly,

$$\left\| \hat{P}(z) - P(z) \right\|_{z=e^{jw}} \leq \left\| \frac{0.15z^4 - 0.138z^3 - 0.13z + 0.117}{z^3(z - 0.9)(z - 0.92)} \Delta \right\|_{z=e^{jw}},$$

so we can define

$$W_1 = \frac{0.15z^4 - 0.138z^3 - 0.13z + 0.117}{z^3(z - 0.9)(z - 0.92)}.$$

We can see that the conditions for Proposition 6 are satisified. With this P in place, it can be shown that the fast single-rate system (Figure 9.2) is closed-loop unstable, while the dual-rate inferential control system (Figure 9.4) with $n = 4$ is closed-loop stable. This is shown in the closed-loop step responses in Figures 9.9 and 9.10. This shows that the dual-rate inferential system is indeed more robust than the corresponding fast single-rate system!

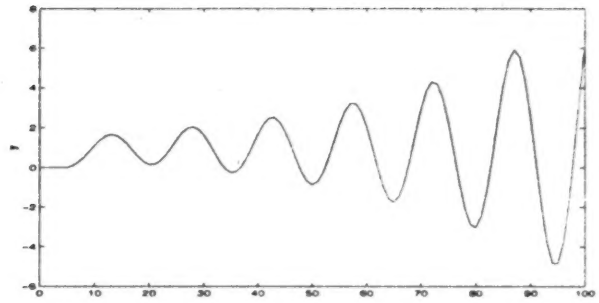


Figure 9.9: *Step response in y for the fast single-rate system.*

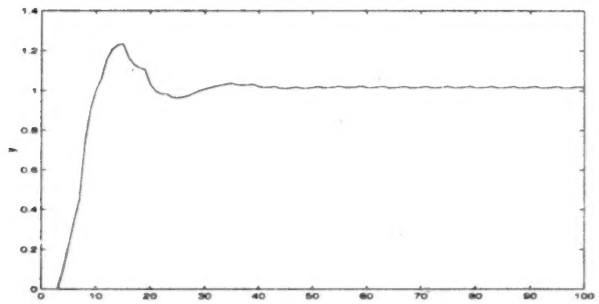


Figure 9.10: *Step response in y for the dual-rate inferential control system.*

9.7 Conclusion

In the preceding sections, we showed that in the absence of model-plant mismatch, the multirate inferential control system is as stable as the fast single-rate system, and the multirate inferential control system can recover the performance of the fast single-rate system. We also studied the stability robustness of the dual-rate inferential control system assuming an additive uncertainty model. We point out here that similar result holds if a multiplicative uncertainty model is used; in this case P belongs to the class

$$\{(I + \Delta W_1)\hat{P} : \quad \|\Delta\|_\infty < 1\}.$$

Similar to the condition in (9.8) in Proposition 6, the robust stability condition for this case is

$$\|\underline{W}_1\underline{\hat{P}}(I + \underline{K}\,\underline{\hat{P}})^{-1}\underline{K}R_1\|_\infty < 1.$$

Based on this, we can make the same conclusion that the dual-rate inferential control scheme is advantageous in stability robustness over the fast single-rate control scheme.

Chapter 10

Conclusions and Future Work

10.1 Conclusions

The main contributions of this thesis are:

- A tutorial introduction to three popular subspace based identification schemes. All the three algorithms are illustrated with experimental and simulation examples.
- An overview of the multirate control schemes.
- Development of a two-step multirate identification scheme based on the following steps (1) estimation of the lifted model and (2) computation of the fast sampled model from the lifted model. The unique problems in multirate identification, for example, the accuracy of time-delay estimation, the observability in the presence of time-delay and the causality problem, are discussed in detail. This multirate identification scheme has been applied to SISO multirate systems where the ratio between the output sampling interval and control interval is a rational number; and to MIMO multirate systems where all the control intervals are uniform, all the output sampling intervals are distinct but are integer multiples of the control interval. The subspace-based multirate identification schemes are demonstrated with both experimental and simulation examples.
- Estimation of the fast sampled transfer function model from multirate data: Frequency analysis of the multirate identification in the polynomial domain demonstrates the importance of the design of the excitation sequence. Industrial application shows that the proposed scheme works well even with a relatively small data set.

- Presentation of a multirate closed-loop identification scheme: A multirate closed-loop identification algorithm which can estimate the fast-sampled process model in either one step or two steps is developed. This identification scheme can handle the most general class of processes.

- A multirate inferential model-based predictive control scheme is proposed. This control scheme is discussed in both the polynomial and state-space domains. An experimental example is explored to evaluate the model-based multirate inferential control scheme.

- Analysis of the nominal performance and robust stability of the multirate inferential control systems. It can be proved theoretically that the multirate inferential control system may lose a little in performance but at the same time gain in stability robustness. This result is validated by a simulation example.

This thesis proposed a practical way to estimate fast-sampled process models from multirate data, and hence made it possible to control the industrial multirate processes effectively with the model-based inferential control schemes.

10.2 Future Work

The following problems are interesting, important, and worthy of furthur investigation:

- **Validation of estimated fast-sampled models**: In the experimental and simulation examples, we used the slowly sampled output to estimate the fast-sampled model, and the fast-sampled outputs to validate the fast-sampled model. But in industry, the fast-sampled output measurement is not available in multirate systems. It is challenging and of great importance to develop a method to validate the fast-sampled model only based on the slowly sampled output measurements.

- **Analysis of subspace-based identification algorithms in the frequency domain**: Subspace based identification schemes are natural choices for modeling multi-input and multi-output systems which are common in chemical industries. Providing another dimension of insight into the processes, analysis of identification schemes in the frequency domain is important and practical. So far, little work has been done in analyzing subspace-based identification schemes in the frequency domain.

- **Constrained system modeling and identification**: Currently, all the identification algorithms, in both the polynomial domain and state-space domain, treat the processes as black boxes. In practice, these schemes require inserting excitation signals into the processes. Excitation is not desired in chemical industries because they degrade the performance of control loops. On the other hand, operational data are abundant in industries. It would be economical to estimate process models from operational data. But unfortunately, it is difficult to do so, since usually operational data does not contain enough excitaion which is necessary for the black-box schemes. Incorporating the knowledge of the operators and engineers about the processes into the modeling schemes, constrained identification schemes provide a new and promising way for estimating dynamic models from operational data. Theoretical analysis of the constrained schemes such as identifiablity, convergence and consistency would provide solid ground for applying constrained identification schemes in industries.

- **Robust stability and performance analysis of multirate inferential control systems**: Model-plant mismatch (MPM) is inevitable in the chemical process industry due to factors such as scaling, fouling, varying catalyst activity (in chemical reactors for example), equipment degradation with time. The stability of a special class of MR inferential control systems in the presence of MPM has been investigated, but the stability of more general MR inferential control systems in the presence of MPM should also be analyzed. This will constitute an extension to the work reported in Chapter 9.

- **Numerical robust stability analysis of identification algorithms**: Many of the identification problems can be solved as optimization problems. Ideally all the identification schemes can give good results, but in practice many schemes can not even give reasonable results because of the existence of process disturbances and measurement noises. Numerical analysis shows that numerical structures of many identification algorithms are sensitive to poor excitation, process disturbances and measurement noises. Research shows that sensitivity of many identification algorithms can be improved by formulating modified optimization problems. Robust stability analysis of the identification schemes is of great importance since process disturbances and measurement noises are unavoidable in practice.

Bibliography

[1] R. Amirthalingam, S.W. Sung, and J.H. Lee. Two-step procedure for data-based modeling for inferential control applications. *Aiche Journal*, pages 1974–1988, 2000.

[2] K.J. Åström. Lectures on the identfication problem - The least square method. *Technical Report 6806, Division of Automatic Control, Lund Institute of Technology, Lund, Sweden*, 1968.

[3] K.J. Åström and T. Bohlin. Numerical identification of linear dynamic systems from normal operating records. *IFAC Symposium on Self-Adaptive Systems, Teddington, England*, 1965.

[4] H.S. Black. Inventing the negative feedback amplifier. *IEEE Spectrum*, pages 55–60, 1977.

[5] H.W. Bode. Feedback - the history of an idea. *Selected Papers on Mathematical Trendsin Control Theory*, pages 106–123, 1964.

[6] G.E.P. Box and G.M. Jenkins. *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco, 1970.

[7] C.B. Brosilow. The structure and design of Smith predictors from the viewpoint of inferential control. *Proc. of Joint Automatic Control Conf.*, 1979.

[8] P.E. Caines. Prediction error identification methods for stationary stochastic processes. *IEEE Trans. Automatic Control*, 21:598–600, 1976.

[9] T. Chen and B.A. Francis. Linear time-varying $\mathcal{H}_2$ optimal control of sampled-data systems. *Automatica*, 27:963–974, 1991.

[10] T. Chen and B.A. Francis. *Optimal Sampled-data Control Systems.* Springer, London, 1995.

[11] T. Chen and L. Qiu. $\mathcal{H}_\infty$ design of general multirate sampled-data control systems. *Automatica*, 30:1139–1152, 1994.

[12] Y.M. Cho and T. Kailath. Fast subspace-based system identification: an instrumental variable approach. *Automatica*, 31:903–905, 1995.

[13] C.T. Chou and M. Verhaegen. Subspace algorithms for the identification of multivariable dynamic errors-in-variables models. *Automatica*, pages 1857–1869, 1997.

[14] C.T. Chou, M. Verhaegen, and R. Johansson. Continuous-time identification of siso systems using laguerre functions. *IEEE Transactions on Signal Processing*, 47:349–362, 1999.

[15] D.W. Clarke, C. Mohtadi, and P.S. Tuffs. Generalized predictive control. *Automatica*, 23:137–160, 1987.

[16] C.R. Cutler and L. Ramaker, B. Dynamic matrix control - A computer control algorithm. *Proc. of Joint Automatic Control Conf.*, San Francisco:California, 1980.

[17] F. M. D'Hulster and A. R. van Cauwenberghe. Application of parameter adaptive inferential control to non-continuously measurable quality variables. *Proceedings of the Second World Congress of Chemical Engineering*, 1981.

[18] J.C. Doyle. Analysis of control systems with structured uncertainty. *IEE Proc. Part D*, 129:242, 1982.

[19] J.C. Doyle. Structured uncertainty in control system design. *Proc. IEEE Conf. on Decision and Control*, Lauderdale, FL, 1985.

[20] J.C. Doyle, B.A. Francis, and A.R. Tannenbaum. *Feedback Control Theory*. Maxwell Macmillan Canada, 1992.

[21] L. Dugard and I.D. Landau. Recursive output error identification algorithms. *Automatica*, 16:443–462, 1980.

[22] U. Forssell and L. Ljung. Closed-loop identification revisited. *Automatica*, 35:1215–1241, 1999.

[23] B.A. Francis and T.T. Georgiou. Stability theory for linear time-invariant plants with periodic digital controllers. *IEEE Transactions on Automatic Control*, 33:820–832, 1988.

[24] B. Friedland. Sampled-data control systems containing periodically varying members. *In Proc. First IFAC Congress*, pages 361–367, 1960.

[25] H. Garnier, P. Sibille, and A. Richard. Continuous-time canonical model identification via Poisson moment functionals. *Proceedings of CDC*, pages 3004–3009, 1995.

[26] S.I. Grossman. *Elementary Linear Algebra*. Wadsworth Publishing Company, 1980.

[27] R.D. Gudi, S.L. Shah, and M.R. Gray. The role of adaptive Kalman filter as a software sensor and its application to a bioreactor. *Proceedings of 12th IFAC World Congress*, pages 221–225, 1993.

[28] R.D. Gudi, S.L. Shah, and M.R. Gray. Mulitrate state and parameter estimation in an antibiotic fermentation with delayed measurements. *Biotechnology and Bioengineering*, 44:1271–1278, 1994.

[29] M.T. Guilandoust, A.J. Morris, and M.T. Tham. Estimation and control of distillation product composition using tray temperature measurements. *DYCORD*, pages 203–208, 1986.

[30] M.T. Guilandoust, A.J. Morris, and M.T. Tham. An adaptive estimation algorithm for inferential control. *Ind. Eng. Chem. Process Des. Dev.*, 27:1658–1664, 1988.

[31] I. Gustavsson, L. Ljung, and T. Soderstrom. Identification of process in closed loop – Identifiability and accuracy aspects. *Automatica*, 13:59–75, 1977.

[32] E.J. Hannan. The asymptotic theory of linear time series models. *J. App. Prob.*, 10:135–145, 1973.

[33] B. Haverkamp, C.T. Chou, M. Verhaegen, and R. Johansson. Identification of continuous-time MIMO state space models from sampled data. *Proceedings of CDC*, pages 1539–1544, 1996.

[34] B. Ho and R.E. Kalman. Efficient contruction of linear state variable models from input/output functions. *Regelungstechnik*, 14:545–548, 1966.

[35] H. Hotelling. Relations between two sets of variables. *Biometrika*, 28:321–377, 1936.

[36] T.C. Hsia. *Identification: Least Squares Methods.* Lexington Books, 1977.

[37] M. Jansson and B. Wahlberg. On consistency of subspace methods for system identification. *Automatica*, 34:1507–1519, 1998.

[38] R.A. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis.* Prentice Hall, 1992.

[39] P.V. Kabaila and G.C. Goodwin. On the estimation of the parameters of an optimal interpolator when the class of interpolater is restricted. *SIAM J. Control and Optimization*, 18(2):121–144, 1980.

[40] M. Kano, K. Miyazaki, S. Hasebe, and I. Hashimoto. Inferential control system of distillation compositions using dynamic partial least squares regression. *Journal of Process Control*, pages 157–166, 2000.

[41] K. Khargonekar, P.P.and Poolla and A. Tannenbaum. Robust control of linear time-invariant plants using periodic compensation. *IEEE Transactions on Automatic Control*, 30:1088–1096, 1985.

[42] P.P. Khargonekar, K. Poolla, and A. Tannenbaum. Robust control of linear time-invariant plants using periodic compensation. *IEEE Transactions on Automatic Control*, 30:1088–1096, 1985.

[43] G.M. Kranc. Input-output analysis of multirate feedback systems. *IEEE Transactions on Automatic Control*, 3:21–28, 1957.

[44] S.Y. Kung. A new identification and model reduction algorithm via singular value decomposition. *Proc. 12th Asilomar Conf. on Circuits, Systems and Computers*, pages 705–714, 1978.

[45] W.E. Larimore. Canonical variate analysis in identification, filtering, and adaptive control. *Proceedings of the 29th Conference on Decision and Control*, pages 596–604, 1990.

[46] C. Leake. *Theory and Problems of Linear Algebra.* Educational Publishers, 1969.

[47] J.H. Lee and M. Morari. Robust inferential control of multi-rate sampled-data systems. *Chemical Engineering Science*, 47:865–885, 1992.

[48] J.H. Lee and Y.H. Park. Optimal production of curdlan by agrobacterium sp with feedback inferential control of optimal ph profile. *Biotechnology Letters*, pages 525–530, 2001.

[49] D. Li, S.L. Shah, and T. Chen. Identification of fast-rate models from multirate data. *International Journal of Control*, pages 680–689, 2001.

[50] D. Li, S.L. Shah, T. Chen, and R. Patwardhan. System identification and long-range predictive control of multi-rate systems. *Proceedings of American Control Conference*, pages 336–340, 1999.

[51] D. Li, S.L. Shah, T. Chen, and K.Z. Qi. Application of dual-rate modeling to CCR octane quality inferential control. *DYCOPS*, pages 417–421, 2001.

[52] L. Ljung. On convergence for prediction error identification methods. *Technical Report 7405, Division of Automatic Control, Lund Institute of Technology, Sweden*, 1974.

[53] L. Ljung. On the consistency of prediction error identification methods. *In System Identification, Advances and Case Studies(R.K.Mehra and D.G.Lainiotis, eds.)*, 1976.

[54] L. Ljung. Convergence of recursive estimators. *5th IFAC Symp. on Identification and System Parameter Estimation*, pages 131–144, 1979.

[55] L. Ljung. *System Identification, Theory for the user.* Prentice Hall, 1999.

[56] L. Ljung and K. Glover. Frequency domain versus time domain methods in system identification. *Automatica*, 17:71–86, 1981.

[57] W.P. Lu and D.G. Fisher. Output estimation with multi-rate sampling. *International Journal of Control*, 48:149–160, 1988.

[58] W.P. Lu and D.G. Fisher. Output estimation with multi-rate sampling. *International Journal of Control*, 48:149–160, 1988.

[59] D.Q. Mayne. A method for estimating discrete time transfer functions. *In Advances in Computer Control, 2nd UKAC Control Convention, University of Bristol*, 1967.

[60] J.M. Mendel. *Discrete Techniques of Parameter Estimation: The Equation Error Formulation.* Marcel Dekker, New York, 1973.

[61] M. Morari and E. Zafiriou. *Robust Process Control.* Prentice Hall, 1989.

[62] H. Nyquist. Regeneration theory. *Bell Syst. Tech. J.*, 11:126–147, 1932.

[63] M. Ohshima, I. Hashimoto, M. Takeda, T. Yoneyama, and F. Goto. Multi-rate multivariable model predictive control and its application to a semi-commercial polymerization reactor. *Proceedings of American Control Conference*, 2:1576-1581, 1992.

[64] R.M. Oisiovici and S.L. Cruz. Inferential control of high-purity multicomponent batch distillation columns using an extended kalman filter. *Industrial and Engineering Chemistry Research*, pages 2628–2639, 2001.

[65] O. Reiersøl. Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica*, 9:1–23, 1941.

[66] M. F. Sagfors and H. T. Toivonen. H_∞ control of multirate sampled-data systems: A state-space approach. *Automatica*, 34:415–428, 1998.

[67] S.L. Shah. A tutorial introduction to contrained long range predictive control. *Pulp and Paper Canada*, 4:57–63, 1996.

[68] T. Söderström and P. Stoica. *System Identification*. Prentice Hall, 1989.

[69] T. Söderström and P.G. Stoica. *Instrumental Variable Methods for System Identification*. Springer, Berlin, 1983.

[70] V. Solo. *Time Series Recursion and Stochastic Approximation*. Thesis, Australian National University, Canberra, 1978.

[71] A.K. Tangirala, D. Li, R. Patwardhan, S.L. Shah, and T. Chen. Issues in multirate process control. *Proceedings of American Control Conference*, pages 2771-2775, 1999.

[72] A.K. Tangirala, D. Li, R. Patwardhan, S.L. Shah, and T. Chen. Issues in multirate process control. *Automatica*, to appear in October, 2001.

[73] B. Unbehauen and G.P. Rao. Continuous-time approaches to system identification–a survey. *Automatica*, 26:23–35, 1990.

[74] P.P. Vaidyanathan. *Multirate Systems and Filter Banks*. Prentice-Hall, 1993.

[75] P. Van Overschee and B. DeMoor. Subspace algorithms for the stochastic identification problem. *Automatica*, 29:649–660, 1993.

[76] P. Van Overschee and B. DeMoor. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30:75–93, 1994.

[77] P. Van Overschee and B. DeMoor. A unifying theorem for three subspace system identification algorithms. *Automatica*, 31:1853–1864, 1995.

[78] M. Verhaegen. Application of a subspace model identification technique to identify lti systems operating in closed-loop. *Automatica*, 29:1027–1040, 1993.

[79] M. Verhaegen. Identification of the deterministic part of MIMO state space models given in innovation form from input-output data. *Automatica*, 30:61–74, 1994.

[80] M. Verhaegen and P. Dewilde. Subspace model identification. Part I: The output-error state space model identification class of algorithms. *International Journal of Control*, 56:1187–1210, 1992.

[81] M. Verhaegen and P. Dewilde. Subspace model identification. Part II: Analysis of the elementary output-error state space model identification algorithm. *International Journal of Control*, 56:1211–1241, 1992.

[82] M. Verhaegen and X. Yu. A class of subspace model identification algorithms to identify periodically and arbitrarily time-varying systems. *Automatica*, 32:201–216, 1995.

[83] M. Viberg. Subspace-based methods for the identification of linear time-invariant systems. *Automatica*, 31:1835–1851, 1995.

[84] S. Voorakaranam and B. Joseph. Model predictive inferential control with application to a composites manufacturing process. *Industrial and Engineering Chemistry Research*, pages 433–450, 1999.

[85] P. Whittle. *Hypothesis Testing in Time Series Analysis*. Thesis, Uppsala University, New York, 1951.

[86] K.Y. Wong and E. Polak. Identification of linear discrete time systems using the instrumental variable approach. *IEEE Transactions on Automatic Control*, 12:707–718, 1967.

[87] D.C. Youla, J.J. Bongiorno, and H.A. Jabr. Modern Wiener-Hopf design of optimal controllers - Part I: The single input-output case. *IEEE Transactions on Automatic Control*, 21:3–13, 1976.

[83] D.C. Youla, H.A. Jabr, and J.J. Bongiorno. Modern Wiener-Hopf design of optimal controllers - Part II: The multivariable case. *IEEE Transactions on Automatic Control*, 21:319–338, 1976.

[89] P.C. Young. On a weighted steepest descent method of process parameter estimation. *Report, Cambridge University, Engineering Laboratory*, 1965.